Topics on estimation, prediction and bounding risk for multivariate extremes

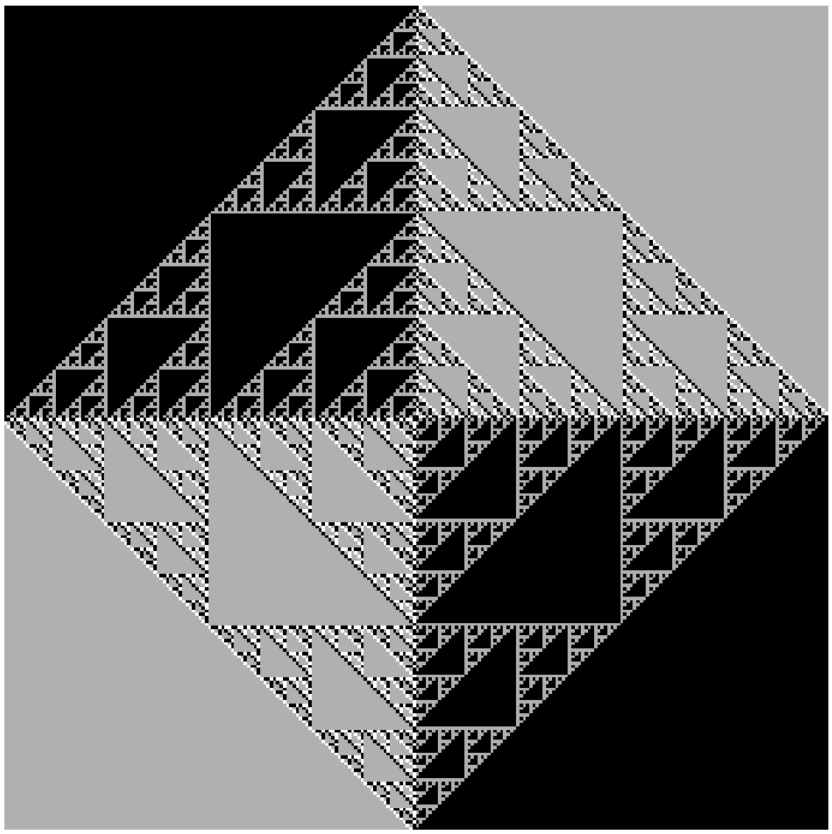
by

Robert Alohimakalani Yuen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2015

Doctoral Committee:

Associate Professor Stilian A. Stoev, Chair
Assistant Professor Veronica Berrocal
Professor Tailen Hsing
Professor Kerby Shedden



Ma ka hana ka ʻike

Knowledge is gained by doing.

For Shannon

ACKNOWLEDGEMENTS

I first want to thank my advisor Stilian Stoev for his encouragement, shrewd advice and most of all his validation of my academic per suits. I also warmly thank professors Tailen Hsing, Kerby Shedden and Veronica Berrocal for serving on my thesis committee. In particular, I want to thank professor Berrocal for her guidance and advice during my time as a graduate student.

I would like to acknowledge the following people who have all played instrumental roles during my time as a student: Joyce Nielson, Brian Mclain, Amy Bailey, Dan Cooley, Tilmann Gneiting, Yizao Wang , Kim Paxton, Daniel Almiral and Peter Guttorp. Generous funding was provided by the Kamehameha Schools Bishop Estate and the Horace H. Rackham Graduate School. I also want to say a special thanks to Judith McDonald and the administrative staff of the University of Michigan Department of Statistics for all their assistance.

Many thanks to my parents, Dominica, Joseph, Anthony and Linda, my grandfather Robert and my sister Rachel. To my son Kaleo, thank you for teaching me that happiness is anywhere you care look. Finally, I want to thank my wife Shannon for her sacrifice, unwavering support and spirited companionship. She deserves much credit for this work and without her this undertaking would not be possible.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES vi

LIST OF TABLES ix

ABSTRACT xi

CHAPTER

1. Introduction 1

1.1 Univariate Extremes 2
1.2 Multivariate Extremes 3
1.3 Regular Variation 7

2. Minimum distance estimation for max-stable models 10

2.1 Examples of max-stable models 12
2.2 Consistency and asymptotic normality 17
2.3 CRPS M-estimation for max-stable models 20
2.4 Simulation 23
2.4.1 Example: multivariate logistic model 24
2.4.2 Example: Schlather model 25
2.4.3 Example: max-linear model 27
2.5 Discussion 28
2.6 Proofs 29

3. Hierarchical Gauss-Pareto models for spatial prediction of extreme precipitation 38

3.1 Pareto processes 40

3.2 A log-Gauss-Pareto model for extreme precipitation 44
3.2.1 Model hierarchy 46
3.2.2 MCMC sampling 50
3.2.3 Spatial prediction 53
3.3 Extreme Summer precipitation over southern Sweden 55
3.4 Discussion 61

4. Tawn-Molchanov random vectors and bounding Value-at-Risk for the maximum loss 64

4.1 The Tawn-Molchanov model 67
4.2 The Tawn–Molchanov upper bound on VaR–max 72
4.2.1 Example: VaR-max for simulated max-linear model 76
4.3 Estimation for the Tawn-Molchanov model 78
4.4 An application to Industry portfolios 84
4.4.1 Estimating upper bounds on VaR-max for portfolio losses 89
4.5 Scaling VaR–max of block maxima to daily VaR–max 91
4.6 Discussion 96
4.7 Notes on the quadratic program 97
4.8 Notes on the TM model and k-way extremal dependence . . 99
4.9 Proofs 101

5. Bounding Value-at-Risk for the sum of dependent losses . . . 104

5.1 Introduction 104
5.2 Linear semi-infinite programming 109
5.3 Main results 112
5.4 Examples 119
5.4.1 Example: bi-variate constraints 119
5.4.2 Example: single d-variate constraint 123
5.5 Discussion 129
5.6 Conic programming. 131
5.7 Strong duality, optimality and reducibility of SIPs 133
5.8 Proofs 135

APPENDIX 141
A.1 Code for fitting max-linear models via CRPS 142
A.2 Code for fitting a Tawn-Molchanov max-stable model 144

BIBLIOGRAPHY 147

LIST OF FIGURES

Figure

2.1 Schlather max-stable model realizations using correlation functions of Table 2.1 under varying parameter settings. **Top:** Stable correlation function. **Middle:** Matérn correlation function. **Bottom:** Cauchy correlation function. Realizations were generated using the `R` package `SpatialExtremes` (Ribatet, 2011). The circles indicate locations of "observation staions" in the simulation study of Section 2.4. 16

3.1 Four realizations from the log-Gauss-Pareto process (3.6) with $Z \sim \text{GPD}(\mu = 0, \sigma = 1, \xi = 0.5)$, $\lambda \sim \exp(1)$, $\omega \sim \text{uniform}(S)$, and $\alpha = 0.5$. The process has been censored below 0.1mm. The ▲ correspond to the process origin ω for each of the four realizations. 47

3.2 Map of synoptic stations over south central Sweden. 56

3.3 Deviance scores (3.10) of model fit versus smoothness parameter α. 58

3.4 Posterior mean of origin centers $\hat{\omega}_i$ versus location of maximum observation $s_i^{\max}$ 59

3.5 Distribution of posterior means for the range λ_i and intensities Z_i for each fitted date $i = 1, \ldots, 59$. 60

3.6 Probability integral transform histograms of the predictive distributions at 21 validation sites . Solid horizontal lines correspond to perfect uniformity. Confidence bands (dashed lines) are provided for reference, bin heights from random draws of a standard uniform distribution should fall outside of the dashed lines in approximately ten percent of cases. 61

4.1 (a) Interpreting $\boldsymbol{\beta}$ for a 3 asset portfolio: Since $\beta_{\{1\}} = 0.5$, half of the shocks to asset 1 occur independently from 2 and 3 and since $\beta_{\{1,2,3\}} = 0.5$, the remaining half of the shocks affecting 1 also affect assets 2 and 3. Similarly, since $\beta_{\{2\}} = 0.25$, one quarter of the shocks to asset 2 occur independently of 1 and 3; since $\beta_{\{2,3\}} = 0.25$, another quarter of the shocks to asset 2 occur independently of 1 but also affect 3, and since $\beta_{\{1,2,3\}} = 0.5$ the remaining half of the shocks affecting 2 also affect the entire portfolio. Finally, the sum of all β_{J_k} is equal to the extremal coefficient $\vartheta_{\boldsymbol{X}}(\{1,\ldots,d\}) = 1.75$. (b) Pairwise tail-dependence graph. Edge weights between nodes correspond to bivariate tail dependence coefficients $w(i,j)$ from (4.13). Edge weights at each node indicate level of asymptotic independence $\beta_{\{j\}}$ for components $j \in \{1,2,3\}$. 69

4.2 Ratio of Value-at-Risk for component-wise maxima of max-linear model $\boldsymbol{V}$ versus TM model $\boldsymbol{W} = \mathrm{TM}(\boldsymbol{V})$. Grey lines indicate 100 replicates of the ratio (4.25) and the dark line reports the median. Rows correspond to coefficient matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ in (4.24) for max-linear model (4.23). Columns indicate different values of the shape and scale parameters $\boldsymbol{\xi}$ and $\boldsymbol{\sigma}$ for the margins of $\boldsymbol{V}$. 79

4.3 TM estimation results for max-linear model $\boldsymbol{A}_1$. Reported are boxplots of 1000 replicates of fitted TM model coefficients $\hat{\beta}_{J_k}$ using sample size $n = 50$. The true β_{J_k} for this model are indicated by the dark points •. The top panel corresponds to lighter tails ($\xi = 0.5$) vs heavier tails ($\xi = 1$) in the bottom panel. 85

4.4 TM estimation results for max-linear model $\boldsymbol{A}_2$. Reported are boxplots of 1000 replicates of fitted TM model coefficients $\hat{\beta}_{J_k}$ using sample size $n = 50$. The true β_{J_k} for this model are indicated by the dark points •. The top panel corresponds to lighter tails ($\xi = 0.5$) vs heavier tails ($\xi = 1$) in the bottom panel. 86

4.5 Asset-wise annual maximum of negative daily returns for 5 industry portfolio. **Top:** Original scale. **Bottom:** Fréchet scale. 88

4.6 Illustration of the TM model applied to a 5-dimensional portfolio of stock market sectors. 89

4.7 Backtesting time series for annual block maxima. Solid line indicates maxima $\{\boldsymbol{V}_i^{\vee}\}_{i=1}^n$ of the 5 industry portfolio in Figure 4.5. Broken lines correspond to the series of estimated $\mathrm{VaR}_\alpha(\boldsymbol{V}^{\vee})$ at $\alpha \in \{.900, .950, .975, .990\}$ used in the backtest. 92

4.8 VaR-max ratios for daily maxima $\boldsymbol{R}^{\vee}$ versus scaled TM model $M^{-\xi^*}\boldsymbol{W}^{\vee}$ using block size $M \in \{20, 60, 250\}$. Grey lines indicate 100 empirical versions of the ratio $\mathrm{VaR}_\alpha(\boldsymbol{R}^{\vee})/\mathrm{VaR}_\alpha(M^{-\xi^*}\boldsymbol{W}^{\vee})$ and the dark line reports the median. Columns indicate different values of the shape and scale parameters for the margins of $\boldsymbol{W}$. 96

5.1 Results of two experiments. **left:** $d = 6$. **right:** $d = 8$. 100 random discrete spectral measures (H_{true}) were drawn and ordered along the x-axis according to their value of $\rho(H_{\text{true}}, \xi)$. Bounds $(\inf_H, \sup_H)$ correspond to optimization over all spectral measures with identical margins and fixed bivariate extremal coefficients. 122
5.2 Upper and lower bounds on extreme VaR ρ^{ξ} when given a single fixed d-vivariate extremal coefficient ϑ_D constraint. 130

LIST OF TABLES

Table

2.1 Correlation functions for Gaussian random fields. For the Matérn covariance function, B_α is the modified Bessel function of the second kind. 15

2.2 Logistic model simulation results using 500 replications of sample size $n = 100$ and $n = 1000$. Reported are bias and root mean squared error (RMSE) for CRPS and MLE estimates. Coverages are based on plug-in estimates of 95% asymptotic confidence intervals. In the case of the CRPS estimates, confidence intervals are based on (2.14) and computed using the expressions from Corollary 2.9. 25

2.3 CRPS and MCLE estimates for Schlather model. Reported are bias and root mean squared error (RMSE) of 500 replications using sample size $n \in \{100, 1000, 5000\}$. CRPS based confidence intervals for $\boldsymbol{\theta}_0 = (100, 1)$ were calculated using plug-in estimates for the expressions in Corollary 2.9 and resulting 95% coverages are reported. Coverages for MCLE estimates are based on the Godambe information (see Padoan et al., 2010). 26

2.4 Simulation results of the four dimensional two factor max-linear model using 500 replications of sample size $n = 5000$. We compare the CRPS estimator with the M-estimator (M-est) of Einmahl et al. (2012). Their estimator depends on a threshold parameter $\kappa \in (0, n)$, thus reported values are favorable ranges based off of the graphs in Figure 5 of Einmahl et al. (2012) which plot bias and root mean squared error (RMSE) over a wide range of $\kappa \in [40, 1000]$. 28

3.1 Specification of hyper-parameters for MCMC. 57

4.1 GEV parameter settings for VaR-max simulation studies 78

4.2 Medians and inter-quartile range (IQR), in parentheses, for 1000 marginal GEV parameter estimates fitted to samples of size $n = 50$ generated from the max-linear model (4.23) with coefficient matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ of (4.24). Reported are results for the first component V_1 only. Results for the remaining components are nearly identical. . . 84

4.3 TM model backtesting results for 5 industry portfolio. Empirical coverage rates ($\hat{\alpha}$) and naive binomial standard errors for Monthly, Quarterly and Annual TM bounds on VaR-max. The training length and number of tests are given by m and $(n-m)$ respectively. . . . 93
5.1 VaR information $\mathcal{I}(\boldsymbol{\vartheta})$ for all pairwise extremal coefficients. 122
5.2 VaR information $\mathcal{I}(\boldsymbol{\vartheta})$ for a single d-variate extremal coefficient. . . 131

ABSTRACT

Topics on estimation, prediction and bounding risk for multivariate extremes

by

Robert Alohimakalani Yuen

Chair: Stilian A. Stoev

This dissertation consists of results in estimation, prediction and bounding risk for multivariate extremes. Regarding estimation, we establish a consistent and asymptotically normal M-estimator that is applicable to a wide variety of max-stable models, i.e. the class of distributions arising as the limit of component-wise maxima. Such processes play a fundamental role in modeling extreme phenomena, but are challenging to work with due to a lack of tractable likelihoods. Our method circumvents intractable likelihoods, working directly with distribution functions of max-stable processes which are readily available or can be approximated in a precise manner. Our second contribution is in the area of prediction for spatial extremes, specifically extreme precipitation. We introduce Gauss-Pareto random fields as a flexible class of models that capture essential non-trivial extremal dependence characteristics, yet remain amenable to standard Bayesian MCMC techniques. We apply Gauss-Pareto processes to spatial prediction of extreme precipitation over Sweden and show that Gauss-Pareto models yield skillful predictions in practice. Lastly we establish universal bounds on the extreme Value-at-Risk of various functionals of portfolio losses. Specifically, the maximum portfolio loss and the sum of tail dependent losses un-

der given summary measures of tail dependence called extremal coefficients. While extremal coefficients are finite dimensional and consistent estimators are readily obtainable, they do not fully characterize tail dependence. Prior to this work, it was not known how extremal coefficients constrain Value-at-Risk for extreme losses. The solution involves solving an optimization problem over an infinite dimensional space of measures. Here we prove that the optimization problem can be reduced to a convex optimization problem in finite dimensions and develop algorithms to compute the bounds.

CHAPTER 1

Introduction

Dependence in the tails of a multivariate distribution can be quite different than the dependence structure near its 'center'. This phenomena can be observed empirically in data from a variety of fields including insurance, finance, and atmospheric sciences. Probability models that do not account for such tail dependence are theoretically blind of contagion effects present during times of extreme shock. A well known example is the class of Gaussian random vectors which are asymptotically independent unless their distribution is degenerate, i.e. two or more variables are exactly co-linear. This has motivated a large body of research in multivariate extreme value theory (MEVT), in order to characterize classes of probability models with non-trivial tail dependence - i.e. multivariate *extreme value distributions*. While a variety of models have been theoretically established, challenges remain in applications because most often there is a lack tractable likelihoods and conditional distributions are quite challenging to work with (see e.g. Wang and Stoev, 2011; Dombry et al., 2012). These difficulties have hampered statistical inference for multivariate extremes. In particular, estimation and prediction for extreme value processes have remained an open area of research.

This dissertation presents three contributions in the area of statistical inference for multivariate extremes, respectively: *estimation*, *prediction* and *bounding of risk*.

After an introduction to extreme value theory, Chapter 2 introduces a minimum distance estimator for *max-stable models* - the class of distributions arising as the limit of component-wise maxima. Chapter 3 describes a method for prediction of spatial processes given nearby observed extremes. Chapter 4 introduces the Tawn-Molchanov model as a method for bounding extreme Value-at-Risk for the *maximum* of dependent losses. Lastly, Chapter 5 is an extension of concepts from the previous Chapter, providing theoretical bounds on extreme Value-at-Risk for the more common *sum* of dependent losses under fixed extremal coefficients.

1.1 Univariate Extremes

Let $\eta, \eta_1, \eta_2, \ldots$ be independent, identically distributed random variables. If there exists sequences $\{a_n\} > 0$, and $\{b_n\}$, $n = 1, 2, \ldots$ such that

$$a_n^{-1}\{\max_{i \leq n} \eta_i - b_n\} \xrightarrow{d} \zeta, \text{ as } n \to \infty, \tag{1.1}$$

for some non-degenerate random variable ζ, then by the classical results of Fisher and Tippett (1928) and Gnedenko (1941), ζ must be generalized extreme value distributed (GEV), which has the three parameter distribution function

$$G_{\mu,\sigma,\xi}(z) := \exp\Big\{ -(1 + \xi(z-\mu)/\sigma)_+^{-1/\xi} \Big\}, \ \sigma > 0, \tag{1.2}$$

where $z_+ = \max\{z, 0\}$, and where μ, σ and ξ are the location, scale and shape parameters. The cases $\xi > 0, \xi < 0$, and $\xi \to 0$ correspond to Fréchet, reverse Weibull, and Gumbel, distributions respectively. When relation (1.1) holds it is said that η belongs to the *max-domain of attraction* of ζ, denoted $\eta \in \text{MDA}(\zeta)$. (see, e.g. Ch.3 and 6.3 in Embrechts et al., 1997 for a detailed treatment).

The limiting random variable ζ must be *max-stable*, i.e. there must exist sequences

$\{c_n\} > 0$, and $\{d_n\}$, $n = 1, 2, \ldots$ such that for all n

$$c_n^{-1}\{\max_{i\leq n} \zeta_i - d_n\} \stackrel{d}{=} \zeta, \tag{1.3}$$

where $\zeta_1, \zeta_2, \ldots$ are independent copies of ζ. In view of (1.1) - (1.3) we have for large enough n and z, the approximation

$$\mathbb{P}\big(\max_{i\leq n} \eta_i \leq z\big) \approx G^{1/n}_{\mu,\sigma,\xi}\left(c_n^{-1}\{z - d_n\}\right) \approx G_{\mu,\sigma,\xi}(z)$$

for some set of parameters μ, σ, ξ. Hence, the distribution of block maxima $\max_{i\leq n} \eta_i$ can be seen as approximately GEV.

Remark 1.1. More generally, when $\eta_1, \eta_2, \ldots$ is a stationary time series satisfying mild dependence conditions (Leadbetter et al., 1983), then extreme values tend to appear in clusters. In this case, the *extremal index* $\theta \in (0, 1]$ determines the mean cluster size $1/\theta$ and block maxima then follow approximately $G^{\theta}_{\mu,\sigma,\xi}(z)$.

Alternatively, one can model extremes for η by considering excursions over a high threshold u. In this case if $\eta \in \mathrm{MDA}(\zeta)$ with $\zeta \sim G_{\mu,\sigma,\xi}$, then the following relation holds

$$\mathbb{P}(\eta > z + u|\eta > u) \approx \frac{1 - G_{\mu,\sigma,\xi}(z + u)}{1 - G_{\mu,\sigma,\xi}(u)} \approx (1 + \xi z/\tilde{\sigma})^{-1/\xi}_{+}, \tag{1.4}$$

where $\tilde{\sigma} = \sigma + \xi(u - \mu)$. The distribution function $(1 + \xi z/\tilde{\sigma})^{-1/\xi}_{+}$ in (1.4) is called the *generalized Pareto distribution* (GPD) and is the only possible non-degenerate limit of $\mathbb{P}(\eta > z + u|\eta > u)$ as $u \to \infty$. Thus the GPD serves as a model for peaks over thresholds, in same melody as the GEV with respect to block maxima.

1.2 Multivariate Extremes

Consider $\eta_i = \{\eta_i(s)\}_{s\in S},\ i = 1, 2, \ldots$ as independent and identically distributed realizations of certain physical or economic processes. For example, the $\eta_i(s)$ may

model wave-height or pollutant concentration levels at a location s in a spatial region $S \subset \mathbb{R}^2$, or $\eta_i(s)$'s may model returns for a fund s in a portfolio $S \subset \mathbb{N}$. If one is interested in extremes, it is natural to consider the asymptotic behavior of the point-wise maxima. Suppose that, for some $a_n(s) > 0$ and $b_n(s) \in \mathbb{R}$, we have

$$\Big\{\frac{1}{a_n(s)} \max_{i=1,\dots,n} \eta_i(s) - b_n(s)\Big\}_{s\in S} \overset{f.d.d.}{\to} \{\zeta(s)\}_{s\in S}, \quad \text{as } n \to \infty, \tag{1.5}$$

for some non-trivial limit process ζ, where $\overset{f.d.d.}{\to}$ denotes convergence of the finite-dimensional distributions. As with the univariate case, whenever (1.5) holds, we denote $\eta \in \mathrm{MDA}(\zeta)$. The class of *extreme value* processes $\zeta = \{\zeta(s)\}_{s\in S}$ arising in the limit describe the statistical dependence of 'worst case scenaria' and are therefore natural models of multivariate extremes. The limit ζ in (1.5) is necessarily a *max-stable process* in the sense that for all n, there exist $c_n(s) > 0$ and $d_n(s) \in \mathbb{R}$, such that

$$\Big\{\frac{1}{c_n(s)} \max_{i=1,\dots,n} \zeta_i(s) - d_n(s)\Big\}_{s\in S} \overset{f.d.d.}{=} \{\zeta(s)\}_{s\in S},$$

where $\{\zeta_i(s)\}_{s\in S}$ are independent copies of ζ and where $\overset{f.d.d.}{=}$ means equality of all finite-dimensional distributions. In Section 1.3 we elaborate further on the characterization of max-domain of attraction in terms of *multivariate regular variation* (See e.g. Ch.5 of Resnick, 1987).

The dependence structure of the limiting extreme value process ζ rather than its marginals is of utmost interest in practice. Arguably, the type of the marginals is unrelated to the dependence structure of ζ and as it is customarily done, we shall assume that the limit process ζ has been transformed to 1-Fréchet marginals. That is,

$$\mathbb{P}(\zeta(s) \le z) = G_{\sigma(s),\sigma(s),1}(z) = e^{-\sigma(s)/z}, \ z > 0, \tag{1.6}$$

for some scale $\sigma(s) > 0$ (Ch.5 of Resnick, 1987).

Remark 1.2. It is often the case to simplify (1.6) even further and take the marginals of the limit process ζ to be *standard Fréchet*, i.e. $\mathbb{P}(\zeta(s) \leq z) = G_{1,1,1}(z) = e^{-1/z}$ for all $s \in S$. In this case ζ is called *simple max-stable* (SMS).

Let $\zeta = \{\zeta(s)\}_{s\in S}$ be a max-stable process with 1-Fréchet marginals as in (1.6). Then, its finite-dimensional distributions are multivariate max-stable random vectors and they have the following representation:

$$\mathbb{P}(\zeta(s_j) \leq z_i,\ i = 1, \ldots, d) = \exp\Big\{ - \int_{\mathbb{S}_+^{d-1}} \Big(\max_{i=1,\ldots,d} u_i/z_i \Big) H(d\boldsymbol{u}) \Big\}, \tag{1.7}$$

where $z_j > 0,\ s_j \in S,\ j = 1, \ldots, d$ and where $H = H_{s_1,\ldots,s_d}$ is a finite measure on the positive unit sphere

$$\mathbb{S}_+^{d-1} = \{\boldsymbol{u} = (u_j)_{j=1}^d \,:\, u_j \geq 0,\ \sum_{j=1}^d u_j = 1\}$$

known as the *spectral measure* of the max-stable random vector $(\zeta(s_j))_{j=1}^d$ (see e.g. Proposition 5.11 in Resnick, 1987). The integral in the expression (1.7) is referred to as the *tail dependence function* of the max-stable law. In the following Chapter, we will often use the notation:

$$\sigma(\boldsymbol{z}) \equiv \sigma_{s_1,\ldots,s_d}(\boldsymbol{z}) := -\log \mathbb{P}(\zeta(s_j) \leq z_i,\ j = 1, \ldots, d),$$

where $\boldsymbol{z} = (s_j)_{j=1}^d \in \mathbb{R}_+^d$, for the tail dependence function of the max-stable random vector $(\zeta(s_j))_{j=1}^d$.

It readily follows from (1.7) that for all $a_j \geq 0, j = 1, \ldots, d$, the max-linear combination

$$\max_{j=1,\ldots,d} a_j \zeta(s_j), \tag{1.8}$$

is a 1-Fréchet random variable with scale $\int_{\mathbb{S}_+^{d-1}} (\max_{j=1,\ldots,d} a_j u_j) H(d\boldsymbol{u})$. Conversely,

a random vector $(\zeta(s_j))_{j=1}^d$ with the property that all its non-negative max-linear combinations are 1-Fréchet is necessarily multivariate max-stable (de Haan, 1978). This invariance to max-linear combinations is an important feature that will be used in our estimation methodology of Chaper 2

Some max-stable models are readily expressed in terms of their spectral measures while others via tail dependence functions. These representations however are not convenient for computer simulation or in the case of random processes, where one needs a handle on all finite-dimensional distributions. The most common constructive representation of max-stable process models is based on Poisson point processes (de Haan, 1984; Schlather, 2002; Kabluchko et al., 2009). See also Stoev and Taqqu (2005) for an alternative.

Indeed, consider a measure space $(\Omega, \mathcal{F}, \nu)$ and let $\Pi := \{(\Gamma_i, W_i)\}_{i\in\mathbb{N}}$ be a Poisson point process on $\mathbb{R}^+ \times \Omega$ with intensity measure $d\Gamma d\nu$.

Proposition 1.3. *Let $g : S \times \Omega \mapsto [0, \infty)$ be ν-integrable for every $s \in S$ and let*

$$\zeta(s) := \max_{i\in\mathbb{N}} \Gamma_i^{-1} g(s, W_i), \quad s \in S. \tag{1.9}$$

Then, the process $\zeta = \{\zeta(s)\}_{s\in S}$ is max-stable with 1-Fréchet marginals and finite-dimensional distributions:

$$\mathbb{P}(\zeta(s_j) \le z_j,\ j = 1, \dots, d) = \exp\Big\{ -\int_\Omega \Big(\max_{j=1,\dots,d} g(s_j, w)/z_i \Big) \nu(dw) \Big\}. \tag{1.10}$$

Proof. By (1.9), for all $x_j > 0,\ j = 1, \dots, d$,

$$\mathbb{P}(\zeta(s_j) \le z_j,\ i = 1, \dots, d) = \mathbb{P}(\Pi \subset A) = \mathbb{P}(\Pi \cap A^c = \emptyset),$$

where

$$A = \{(\Gamma, w) \in \mathbb{R}_+ \times \Omega\ :\ g(s_j, w)/\Gamma \le z_j,\ j = 1, \dots, d\}.$$

Observe that $A^c = \{(\Gamma, w) : \max_{j=1,\ldots,d} g(s_j, w)/z_j > \Gamma\}$. Since Π is a Poisson point process on $\mathbb{R}_+ \times \Omega$ with intensity $d\Gamma\nu(dw)$,

$$\mathbb{P}(\Pi \cap A^c = \emptyset) = \exp\Big\{ - \int_\Omega \int_0^{\max_{j=1,\ldots,d} g(s_j,w)/z_i} d\Gamma\nu(dw)\Big\},$$

which equals (1.10) and completes the proof. The above argument shows that the integrability of the functions $g(s,\cdot)$ implies the $\zeta(s_j)$'s in (1.9) are non-trivial random variables. □

Relation (1.9) is known as the *de Haan spectral representation* of ζ and $\{g(s,\cdot)\}_{s\in S} \subset L^1_+(\Omega, \mathcal{F}, \nu)$ as the spectral functions of the process. It can be shown that every separable in probability max-stable process has such a representation (see de Haan, 1984 and Proposition 3.2 in Stoev and Taqqu, 2005).

Remark 1.4. The expressions (1.7) and (1.10) may be related through a change of variables (Proposition 5.11 Resnick, 1987). While the spectral measure H in (1.7) is unique, a max-stable process has many different spectral function representations. Nevertheless, relation (1.9) provides a constructive and intuitive representation of ζ, that can be used to build interpretable models.

1.3 Regular Variation

A function $\overline{F}$ is said to be *regularly varying* at ∞ with index $\alpha \in \mathbb{R}$ if for all $x > 0$

$$\lim_{t\to\infty} \frac{\overline{F}(xt)}{\overline{F}(t)} = x^\alpha.$$

When $\alpha = 0$ than we say $\overline{F}$ is *slowly varying.* By considering $\overline{F}(x)/x^\alpha$ it is always possible to write a regularly varying function $\overline{F}$ as

$$\overline{F}(x) = x^\alpha L(x),$$

where $L(x)$ is a slowly varying function. When studying extremes, regular variation arises in the context of a random variable η having regularly varying survival function $\overline{F}(x) = \mathbb{P}(\eta > x)$. More precisely, a non-negative random variable η with survival function $\overline{F}(x)$ is regularly varying if there exists a scalar $\xi \in \mathbb{R}_+$ such that for all $x > 0$,

$$\lim_{t\to\infty} \frac{\overline{F}(xt)}{\overline{F}(t)} = x^{-1/\xi}. \tag{1.11}$$

In fact, the Pareto tail in (1.11) is the only possible non-degenerate limit of the conditional excess $\lim_{t\to\infty} \mathbb{P}(\eta > xt|\eta > t)$ with $x > 1$ (see e.g. Prop. 2.3 of Resnick, 2007). This, in essence, motivates the regular variation framework for univariate extremes. Since the focus of this dissertation is on multivariate extremes, we will mainly be concerned with a natural extension of (1.11) for random vectors in $\boldsymbol{\eta} = (\eta(s_1), \ldots, \eta(s_d) \in \mathbb{R}_+^d$ called *multivariate regular variation.*

Definition 1.5. A non-negative random vector $\boldsymbol{\eta} \in \mathbb{R}_+^d$ is multivariate regularly varying if there exists a Radon measure ν on $\mathbb{R}_+^d \backslash \{\mathbf{0}\}$ called the *exponent measure* such that

$$\lim_{t\to\infty} \frac{\mathbb{P}\left(t^{-1}\boldsymbol{\eta} \in [\mathbf{0}, \boldsymbol{x}]^c\right)}{\mathbb{P}\left(t^{-1}\boldsymbol{\eta} \in [\mathbf{0}, \mathbf{1}]^c\right)} = \nu\left([\mathbf{0}, \boldsymbol{x}]^c\right) \tag{1.12}$$

for all $\boldsymbol{x} \in \mathbb{R}_+^d \backslash \{\mathbf{0}\}$ such that $[\mathbf{0}, \boldsymbol{x}]^c$ is a continuity set of ν.

By Thm 6.1 of Resnick (2007), Definition 1.5 is equivalent to the following spectral measure characterization

Proposition 1.6. *Let $\|\boldsymbol{\eta}\| = \eta(s_1) + \cdots + \eta(s_d)$. A non-negative random vector $\boldsymbol{\eta}$ is multivariate regularly varying if there exists scalars $\rho, \alpha > 0$, a function $b(t) \to \infty$ and a finite measure H on $\mathbb{S}_+^{d-1}$ such that for all $x > 1$*

$$\lim_{t\to\infty} t\mathbb{P}\left(\|\boldsymbol{\eta}\| > xb(t), \frac{\boldsymbol{\eta}}{\|\boldsymbol{\eta}\|} \in A\right) = \rho x^{-\alpha} H(A), \tag{1.13}$$

for any Borel $A \subset \mathbb{S}_+^{d-1}$ such that $H(\partial A) = 0$, i.e. A is a continuity set of H.

It was shown in Balkema and Resnick (1977) (see also Chap. 5 of Resnick, 1987) that regular variation fully characterizes max-domain of attraction in the sense that Relation (1.12) is necessary and sufficient for $\boldsymbol{\eta} \in \mathrm{MDA}(\boldsymbol{\zeta})$ where in fact

$$\mathbb{P}(\boldsymbol{\zeta} \leq \boldsymbol{z}) = \exp\left\{-\nu\left([\mathbf{0}, \boldsymbol{z}]^c\right)\right\}.$$

If the univariate marginals of $\boldsymbol{\zeta}$ are normalized to α-Fréchet, the exponent measure ν, being *homogeneous,* has the property

$$\nu(tA) = t^{-\alpha}\nu(A).$$

Moreover, the limit measure H in (1.13) and the spectral measure of the max-stable $\boldsymbol{\zeta}$ from Relation (1.7) coincide.

CHAPTER 2

Minimum distance estimation for max-stable models

We have seen that max-stable processes form a canonical class of statistical models for multivariate extremes. They appear in a variety of applications ranging from insurance and finance (Embrechts et al., 1997; Finkenstädt and Rootzén, 2004) to spatial extremes such as precipitation (Davison and Blanchet, 2011; Davison et al., 2012) and extreme temperature. Recall that max-stable processes are exactly the class of non-degenerate stochastic processes that arise from limits of independent componentwise maxima. However, most max-stable models suffer from intractable likelihoods, thus prohibiting standard maximum likelihood and Bayesian inference. This has motivated development of maximum *composite* likelihood estimators (MCLE) for max-stable models (Padoan et al., 2010) as well as certain approximate Bayesian approaches (Reich and Shaby, 2012; Erhardt and Smith, 2012).

In contrast to their likelihoods, the cumulative distribution functions (CDFs) for many max-stable models are available in closed form, or they are tractable enough to approximate within arbitrary precision. This motivates statistical inference based on the minimum distance method (Wolfowitz, 1957; Parr and Schucany, 1980). In this Chapter, we propose an M-estimator for parametric max-stable models based on

minimizing distances of the type

$$\int_{\mathbb{R}^d} \left(F_{\boldsymbol{\theta}}\left(\boldsymbol{x}\right) - F_n\left(\boldsymbol{x}\right)\right)^2 \omega\left(d\boldsymbol{x}\right). \tag{2.1}$$

where $F_{\boldsymbol{\theta}}$ is a d-dimensional CDF of a parametric model, F_n is a corresponding empirical CDF and ω is a weighting measure that emphasizes various regions of the sample space $\mathbb{R}^d$. Using elementary manipulations it can be shown that minimizing distances of the type (2.1) is equivalent to minimizing the *continuous ranked probability score* (CRPS) (Gneiting and Raftery, 2007; Szekely and Rizzo, 2005).

Definition 2.1. (CRPS M-estimator) Let ω be a measure that can be tuned to emphasize regions of a sample space $\mathbb{R}^d$. Define the CRPS functional

$$\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{x}) = \int_{\mathbb{R}^d} \left(F_{\boldsymbol{\theta}}\left(\boldsymbol{y}\right) - \mathbf{1}_{\{\boldsymbol{x} \leq \boldsymbol{y}\}}\right)^2 \omega\left(d\boldsymbol{y}\right) \tag{2.2}$$

Then for independent random vectors $\{\boldsymbol{X}_i\}_{i=1}^n$ with common distribution function $F_{\boldsymbol{\theta}_0}$ we define the following CRPS M-estimator for $\boldsymbol{\theta}_0$.

$$\hat{\boldsymbol{\theta}}_n = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \sum_{i=1}^{n} \mathcal{E}_{\boldsymbol{\theta}}\big(\boldsymbol{X}_i\big). \tag{2.3}$$

For simplicity, we shall assume that the parameter space Θ is a compact subset of $\mathbb{R}^p$, for some integer p.

The remainder of this Chapter is organized as follows. In Section 2.1 we introduce several examples of popular max-stable models. In Section 2.2 we establish regularity conditions for consistency and asymptotic normality of the CRPS M-estimator and provide general formulae for calculating its asymptotic covariance matrix. In Section 2.3 we specialize these calculations to the max-stable setting. In Section 2.4 we conduct a simulation study to evaluate the proposed estimator for popular max-stable models.

2.1 Examples of max-stable models

Recall from (1.9) the spectral representation of a max-stable random field

$$X(s) := \max_{i\in\mathbb{N}} \Gamma_i^{-1} g(s, W_i), \quad s \in S, \tag{2.4}$$

where $g(s,\cdot) \in L^1(\Omega, \mathcal{F}, \nu)$ and $\{\Gamma_i, W_i\}_{i=1}^{\infty}$ are points of a Poisson point process with intensity $d\Gamma d\nu$. A great variety of max-stable models can be defined by specifying a measure space $(\Omega, \mathcal{F}, \nu)$ and an accompanying family of spectral functions $\{g(s,\cdot)\}_{s\in S}$ or equivalently through a consistent family of spectral measures or tail dependence functions. We review next several popular max-stable models and their basic features.

- *(Multivariate logistic)* Let $\boldsymbol{X} = (X_i)_{i=1}^d$ have the CDF

$$F_{\boldsymbol{\theta}}(\boldsymbol{x}) = e^{-\sigma_{\boldsymbol{\theta}}(\boldsymbol{x})}, \quad \text{where } \sigma_{\boldsymbol{\theta}}(\boldsymbol{x}) = \lambda\Big(\sum_{i=1}^d x_i^{-1/\alpha}\Big)^{\alpha},$$

for $\theta = (\lambda, \alpha) \in (0,\infty) \times [0,1]$. The parameter α controls the degree of dependence, where $\alpha = 1$ corresponds to independence ($\sigma_{\boldsymbol{\theta}}(\boldsymbol{x}) = \lambda \sum_{i=1}^d x_i^{-1}$), while $\alpha \downarrow 0$ to complete dependence ($\sigma_{\boldsymbol{\theta}}(\boldsymbol{x}) = \lambda \max_{i=1,\ldots,d} x_i^{-1}$, interpreted as a limit).

This model is rather simple since the dependence is exchangeable but it provides a useful benchmark for the performance of the CRPS-based estimators since the MLE is easy to obtain in this case (see Table 2.2 below). The recent works of Fougères et al. (2009) and Fougères et al. (2013) develop far-reaching generalizations of multivariate logistic laws by exploiting connections to sum-stable distributions.

- *(Spectrally Gaussian models)* By viewing $(\Omega, \mathcal{F}, \nu)$ as a probability space, in the case $\nu(\Omega) = 1$, the spectral functions $\{g(s,\cdot)\}_{s\in S}$ in (2.4) become a stochastic process. By picking $g(s, W) = h(W(s))$ to be non-negative transformations of a Gaussian process $W \equiv \{W(s)\}_{s\in S}$, one obtains interesting and tractable max-stable models whose dependence structure is governed by the covariance structure of the un-

derlying Gaussian process W. This typically involves choosing a family of parametric covariance functions $\rho_{\boldsymbol{\theta}}(t,s),\ \boldsymbol{\theta} \in \mathbb{R}^p$ which characterize the dependence structure of the underlying Gaussian process W, and therefore the resultant max-stable random field X. We list a few popular covariance functions in Table 2.1. The well known Smith, Schlather, and Brown-Resnick random field models are of this type (Smith, 1990; Schlather, 2002; Brown and Resnick, 1977; Stoev, 2008; Kabluchko et al., 2009).

∘ *(Schlather models)* Let $W \equiv \{W(s)\}_{s\in\mathbb{R}^k}$ be a stationary Gaussian random field with zero mean and let $g(s,W) := W(s) \vee 0$. Then $X(s)$ in (2.4) has the following tail dependence function

$$\sigma(\boldsymbol{x}) = \mathbb{E}_\nu \max_{i=1,\ldots,d} \big\{(W(s) \vee 0)/x_i\big\}, \quad \boldsymbol{x} = (x_i)_{i=1}^d \in \mathbb{R}_+^d, \tag{2.5}$$

where $\mathbb{E}_\nu$ denotes integration with respect to the 'probability' measure ν.

∘ *(Brown-Resnick)* Let $W = \{W(s)\}_{s\in\mathbb{R}^k}$ be a zero mean Gaussian random field with stationary increments. Set $g(s,W) := \exp\{W(s) - v(s)/2\}$, where $v(s) = \mathbb{E}_\nu(W(s)^2)$ is the 'variance' of $W(s)$. The seminal paper of Brown and Resnick (1977) introduced this model with W – the standard Brownian motion and showed that, surprisingly, the resulting max-stable process X in (2.4) is stationary, even though W is not. The cornerstone work of Kabluchko et al. (2009) showed that $X \equiv \{X(s)\}_{s\in\mathbb{R}^k}$ is stationary for a centered Gaussian process W, with stationary increments. The tail dependence function of X in this case is

$$\sigma(\boldsymbol{x}) = \mathbb{E}_\nu \max_{j=1,\ldots,d} \big\{\exp(W(s_j) - v(s_j)/2)/x_i\big\}, \quad \boldsymbol{x} = (x_j)_{j=1}^d \in \mathbb{R}_+^d. \tag{2.6}$$

It can be shown that the Smith model (Smith, 1990) is a special case of a Brown-Resnick model with a degenerate random field $\{W(s)\} \stackrel{d}{=} \{s^\top \widetilde{\boldsymbol{W}}\},\ s \in \mathbb{R}^k,\ k < d$, where $\widetilde{\boldsymbol{W}}$ is a Gaussian random vector taking values in $\mathbb{R}^k$. The above models

can be deemed *spectrally Gaussian* since their tail dependence functions (and hence spectral measures) are expectations of functions of Gaussian laws. One can consider other stochastic process models for the underlying spectral functions $g(s, \cdot)$ and thus arrive at general *doubly stochastic* max-stable processes. We comment briefly on some practical considerations for inference with spectrally Gaussian models.

Remark 2.2. If $\{g(s, W)\}_{s \in \mathbb{R}^k}$ is a stationary process, then the max-stable process $X = \{X(s)\}_{s \in \mathbb{R}^k}$ is also stationary. It is, however, non-ergodic. In particular, the Schlather models are non-ergodic. This is important in applications, since a single observation of the random field X at an expanding grid, may not yield consistent parameter estimates. On the other hand, under general conditions, the Brown-Resnick random fields driven by a Gaussian process with stationary increments are mixing (Kabluchko et al., 2009; Stoev, 2008). Therefore, consistent statistical inference from a single realization of such max-stable random fields is possible.

Remark 2.3. The Poisson point process construction in (2.4) involves a maximum over an infinite number of terms. As a result, computer simulations of spectrally Gaussian max-stable models necessitates truncation to a finite number. In the case of the Brown-Resnick model, the number of terms required to produce a satisfactory representation can be prohibitively large. While studies for Brown-Resnick processes on $S \subset \mathbb{R}$ have appeared (Engelke et al., 2012), general simulation of Brown-Resnick processes on $S \subset \mathbb{R}^2$ and higher dimension is quite challenging (Oesting et al., 2011). For this reason the remaining discussion of spectrally Gaussian max-stable models, including simulation is restricted to the Schlather model.

Figure 2.1 displays realizations from the Schlather model for the different correlation functions given in Table 2.1. Note that these examples are all (spectrally) isotropic in the sense that the correlation $\rho(t, s)$ of the underlying Gaussian process depends only on the distance $h = \|t - s\|$ between locations t and s. This however is not a requirement in general. Figure 2.1 also provides some visual evidence of how

Table 2.1: Correlation functions for Gaussian random fields. For the Matérn covariance function, B_α is the modified Bessel function of the second kind.

	$\rho_{\boldsymbol{\theta}}(t,s)$, $\boldsymbol{\theta}=(\lambda,\alpha)$, $h=\|t-s\|$
Stable	$\exp\big[-(h/\lambda)^\alpha\big] \quad \lambda>0, \alpha\in(0,2]$
Matérn	$\frac{(\sqrt{2\alpha}h/\lambda)^\alpha}{\alpha(\alpha)2^{\alpha-1}} B_\alpha\big(\sqrt{2\alpha}h/\lambda\big) \quad \lambda>0,\alpha>0$
Cauchy	$(1+(h/\lambda)^2)^{-\alpha} \quad \lambda>0,\alpha>0$

the covariance structure and smoothness of W influence the dependence structure of the resultant max-stable random field X.

• *(Max-linear or spectrally discrete models)* Let $A=(a_{ij})_{d\times k}$ be a matrix with non-negative entries and let $Z_j,\ j=1,\dots,k$ be independent standard 1-Fréchet random variables. Define

$$X_i=\max_{j=1,\dots,k} a_{ij}Z_j,\quad i=1,\dots,d. \tag{2.7}$$

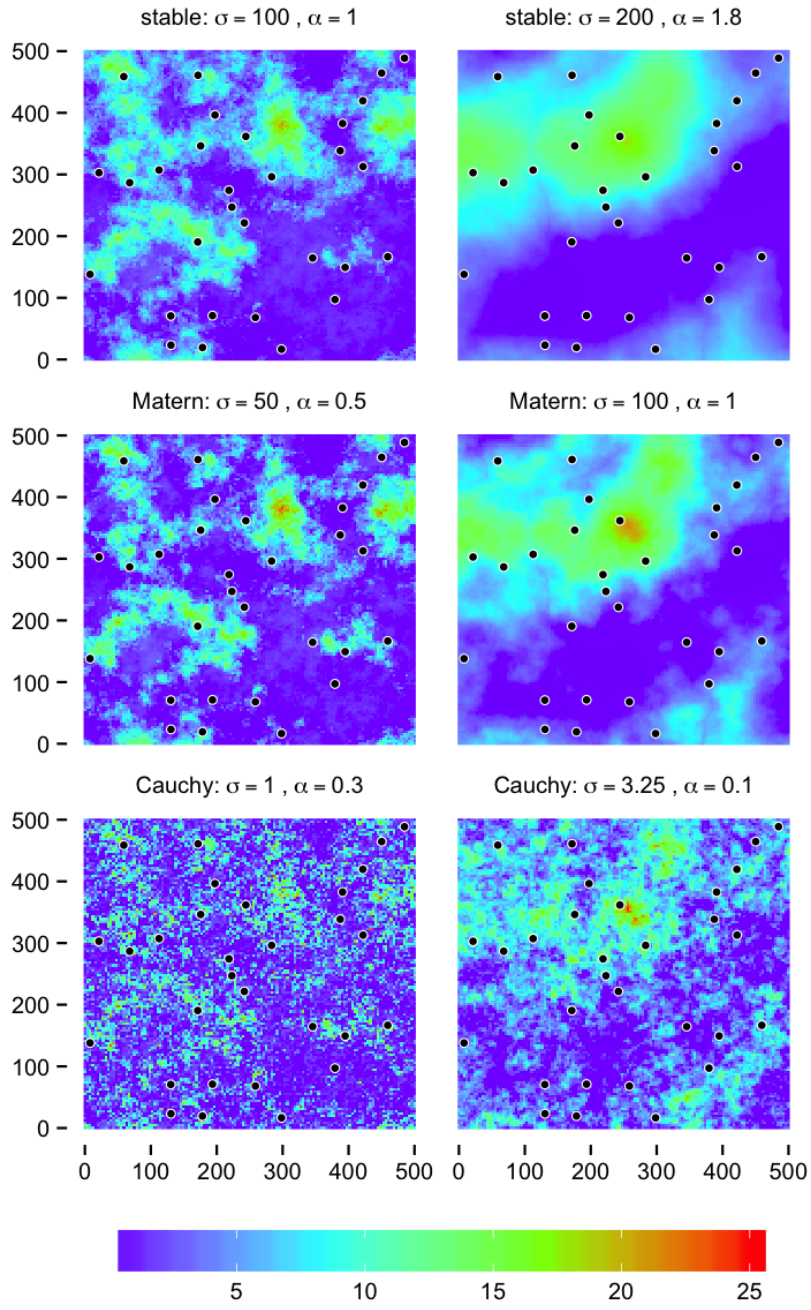
The vector $\boldsymbol{X}=(X_i)_{i=1}^d$ is max-stable. It can be shown that the CDF of $\boldsymbol{X}$ has the form (1.7) were the spectral measure

$$H(dw)=\sum_{j=1}^k |a_{\cdot j}|\delta_{\{a_{\cdot j}/|a_{\cdot j}|\}}(dw), \tag{2.8}$$

is concentrated on the normalized column-vectors of the matrix A, i.e. on $a_{\cdot j}/|a_{\cdot j}| := (a_{ij}/|a_{\cdot j}|)_{i=1}^d$, where $|a_{\cdot j}|=\sum_{i=1}^d a_{ij}$, and where δ_a stands for the Dirac measure with unit mass at the point $a\in\mathbb{R}^d$. Conversely, any max-stable random vector with discrete spectral measure H has a max-linear representation as in (2.7), where the columns of the matrix A may be recovered from (2.8). We shall also call such models spectrally discrete.

Since any spectral measure H can be approximated arbitrarily well by one which is discrete, max-linear models are dense in the class of all max-stable models. As argued

Figure 2.1: Schlather max-stable model realizations using correlation functions of Table 2.1 under varying parameter settings. **Top:** Stable correlation function. **Middle:** Matérn correlation function. **Bottom:** Cauchy correlation function. Realizations were generated using the `R` package `SpatialExtremes` (Ribatet, 2011). The circles indicate locations of "observation staions" in the simulation study of Section 2.4.



in Einmahl et al. (2012), max-linear distributions arise naturally in economics and finance, as models of extreme losses. The Z_j's represent independent shock-factors that lead to various extreme losses in a portfolio $\boldsymbol{X}$ depending on the factor loadings a_{ij}.

Even the bivariate likelihoods for max-linear models are not available in closed form. Consequently, there are limited inference methods for max-linear models. Einmahl et al. (2012) recently introduced an alternative M-estimation methodology. We find max-linear models are particularly well-suited for CRPS-based inference, since their tail dependence function has a simple closed form:

$$\sigma(\boldsymbol{x}) = \sum_{j=1}^{k} \max_{i=1,\dots,d} a_{ij}/x_i, \quad \boldsymbol{x} = (x_i)_{i=1}^{d} \in \mathbb{R}_+^d. \tag{2.9}$$

In Section 2.4 below we provide an example of CRPS-based inference for max-linear models and compare our results with the M-estimator of Einmahl et al. (2012).

2.2 Consistency and asymptotic normality

In this section, we establish general conditions for the consistency and asymptotic normality of CRPS-based M-estimators. This is motivated by questions of inference in max-stable models, but may be of independent interest. Section 2.3 implements and specializes these results to the max-stable setting.

We start with two theorems that are distillations of well known results from the general theory of M-estimators, for example see van der Vaart (1998). Their proofs are given in Section 2.6.

Theorem 2.4. *Let $\boldsymbol{X}, \boldsymbol{X}_1, \boldsymbol{X}_2, \dots$ be iid random vectors with cumulative distribution function $F_{\boldsymbol{\theta}_0}$. Let $\hat{\boldsymbol{\theta}}_n$ be as in Definition 2.1 with $\boldsymbol{\theta}_0$ an interior point of Θ. Suppose that the following conditions hold:*

(i) (identifiability) For all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \Theta$,

$$\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \Rightarrow \omega(\{\boldsymbol{x} \in \mathbb{R}^d : F_{\boldsymbol{\theta}_1}(\boldsymbol{x}) \neq F_{\boldsymbol{\theta}_2}(\boldsymbol{x})\}) > 0 \tag{2.10}$$

(ii) (integrability) For $B(\boldsymbol{\theta}_0) \subset \Theta$, *an open neighborhood of* $\boldsymbol{\theta}_0$

$$\int_{\mathbb{R}^d} \sup_{\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0)} (1 - F_{\boldsymbol{\theta}}(\boldsymbol{x}))\, \omega(d\boldsymbol{x}) < \infty. \tag{2.11}$$

(iii) (continuity) The function $\boldsymbol{\theta} \mapsto \int_{\mathbb{R}^d}(F_{\boldsymbol{\theta}}(\boldsymbol{x}) - F_{\boldsymbol{\theta}_0}(\boldsymbol{x}))^2\omega(d\boldsymbol{x})$ *is continuous in the compact parameter space* $\Theta \subset \mathbb{R}^p$.

Then $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$, *as* $n \to \infty$, *where* $\xrightarrow{p}$ *denotes convergence in probability.*

Theorem 2.5. *Assume the conditions and notation of Theorem 2.4 hold so that in particular,* $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}_0$. *Suppose, moreover, that:*

(i) The measurable function $\boldsymbol{\theta} \mapsto \mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{x})$ *is differentiable at* $\boldsymbol{\theta}_0$ *(for almost every* $\boldsymbol{x}$*) with gradient*

$$\dot{\mathcal{E}}_{\boldsymbol{\theta}_0}(\boldsymbol{x}) := \frac{\partial}{\partial \boldsymbol{\theta}} \mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{x})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}.$$

(ii) There exists a measurable function $L(\boldsymbol{x})$ *with* $\mathbb{E}(L(\boldsymbol{X}))^2 < \infty$, *such that for every* $\boldsymbol{\theta}_1$ *and* $\boldsymbol{\theta}_2$ *in* $B(\boldsymbol{\theta}_0)$

$$|\mathcal{E}_{\boldsymbol{\theta}_1}(\boldsymbol{x}) - \mathcal{E}_{\boldsymbol{\theta}_2}(\boldsymbol{x})| \leq L(\boldsymbol{x})\, \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|. \tag{2.12}$$

(iii) The map $\boldsymbol{\theta} \mapsto \mathbb{E}\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X})$ *admits a second-order Taylor expansion at the point of minimum* $\boldsymbol{\theta}_0$ *with non-singular symmetric second derivative matrix*

$$\mathbf{H}_{\boldsymbol{\theta}_0} := \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \mathbb{E}\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \tag{2.13}$$

Then

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathcal{N}\left(0, \mathbf{H}_{\boldsymbol{\theta}_0}^{-1}\mathbf{J}_{\boldsymbol{\theta}_0}\mathbf{H}_{\boldsymbol{\theta}_0}^{-1}\right), \quad as\ n \to \infty, \tag{2.14}$$

where

$$\mathbf{J}_{\boldsymbol{\theta}_0} := \mathbb{E}\left\{\dot{\mathcal{E}}_{\boldsymbol{\theta}_0}(\boldsymbol{X})\left(\dot{\mathcal{E}}_{\boldsymbol{\theta}_0}(\boldsymbol{X})\right)^\top\right\}. \tag{2.15}$$

The following result provides explicit conditions on the family of CDFs $\{F_{\boldsymbol{\theta}}, \boldsymbol{\theta} \in \boldsymbol{\theta}\}$ that imply conditions *(i)-(iii)* of Theorem 2.5. It also gives concrete expressions for the 'bread' and 'meat' matrices $\mathbf{H}_{\boldsymbol{\theta}_0}$ and $\mathbf{J}_{\boldsymbol{\theta}_0}$ in terms of $F_{\boldsymbol{\theta}}$, which can be used to compute the asymptotic covariances in (2.14). The proof is given in Section 2.6.

Proposition 2.6. *Assume the conditions and notation in Theorem 2.4. Suppose moreover that:*

(i) $\boldsymbol{\theta} \mapsto F_{\boldsymbol{\theta}}(\boldsymbol{y})$ *is twice continuously differentiable for all* $\boldsymbol{\theta}$ *in* $B(\boldsymbol{\theta}_0)$ *with gradient* $\dot{F}_{\boldsymbol{\theta}}(\boldsymbol{y}) := \partial F_{\boldsymbol{\theta}}(\boldsymbol{y})/\partial\boldsymbol{\theta}$ *and second derivative matrix* $\ddot{F}_{\boldsymbol{\theta}}(\boldsymbol{y}) := \partial^2 F_{\boldsymbol{\theta}}(\boldsymbol{y})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top$.

(ii) For all $a \in \mathbb{R}^p$ *with* $\|a\| > 0$

$$\int_{\mathbb{R}^d}\left(a^\top \dot{F}_{\boldsymbol{\theta}_0}(\boldsymbol{y})\right)^2 \omega(d\boldsymbol{y}) > 0. \tag{2.16}$$

(iii) $\int_{\mathbb{R}^d} \sup_{\boldsymbol{\theta}\in B(\boldsymbol{\theta}_0)}\left(\|\dot{F}_{\boldsymbol{\theta}}(\boldsymbol{y})\| + \|\dot{F}_{\boldsymbol{\theta}}(\boldsymbol{y})\|^2 + \|\ddot{F}_{\boldsymbol{\theta}}(\boldsymbol{y})\|\right)\omega(d\boldsymbol{y}) < \infty$.

Then (i)-(iii) of Theorem 2.5 are satisfied and therefore (2.14) *holds, where*

$$\mathbf{H}_{\boldsymbol{\theta}_0} := \int_{\mathbb{R}^d} \dot{F}_{\boldsymbol{\theta}_0}(\boldsymbol{y})\left(\dot{F}_{\boldsymbol{\theta}_0}(\boldsymbol{y})\right)^\top \omega(d\boldsymbol{y}) \tag{2.17}$$

and

$$\mathbf{J}_{\boldsymbol{\theta}_0} := \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \beta_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1, \boldsymbol{y}_2)\, \dot{F}_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1)\left(\dot{F}_{\boldsymbol{\theta}_0}(\boldsymbol{y}_2)\right)^\top \omega(d\boldsymbol{y}_1)\,\omega(d\boldsymbol{y}_2) \tag{2.18}$$

where $\beta_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1, \boldsymbol{y}_2) = F_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1 \wedge \boldsymbol{y}_2) - F_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1)\, F_{\boldsymbol{\theta}_0}(\boldsymbol{y}_2)$.

Remark 2.7. Condition (2.16) ensures that the 'bread' matrix $\mathbf{H}_{\boldsymbol{\theta}_0}$ in (2.17) is non-singular. It is rather mild and fails only if the gradient $\dot{F}_{\boldsymbol{\theta}_0}(\boldsymbol{y})$ lies in a lower dimensional hyper-plane for ω-almost all $\boldsymbol{y}$. In practice, unless the model is over-parameterized this condition typically holds.

Practical inference utilizing the CRPS M-estimator is limited to cases where optimization of $\boldsymbol{\theta} \mapsto \mathcal{E}_{\boldsymbol{\theta}}$ is feasible. Likewise, confidence intervals are only obtained when the matrices $\mathbf{H}_{\boldsymbol{\theta}_0}^{-1}, \mathbf{J}_{\boldsymbol{\theta}_0}$ can be computed. Due to the high dimensionality of the integration involved, this can be difficult under a given weighting measure ω. In the following section we specify a weighting measure that allows efficient computation of the CRPS M-estimator and associated asymptotic covariance matrix under the special case of max-stable models.

2.3 CRPS M-estimation for max-stable models

Our goal is to implement the general CRPS method of the previous section in the case of multivariate max-stable models described in Section 2.1. Here we always have $F_{\boldsymbol{\theta}_0}(\boldsymbol{x}) = \exp\left(-\sigma_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right)$, so in light of Propostion 2.6, this is primarily a matter of specifying the weighing measure ω in the definition of the CRPS (2.2). Overall, the choice of weighting measure is a difficult problem. Indeed, specifying a measure that is 'optimal' in terms of asymptotic efficiency requires knowledge of the unknown parameter $\boldsymbol{\theta}_0$ and moreover, may not be computationally feasible. A complete analysis on the specification of the weighting measure ω is not considered (see discussion Section 4.6). Here, we will choose a specific weighting measure ω^* made explicit by the following polar coordinate transformation

$$\boldsymbol{y} = r\boldsymbol{u} : \quad r = \sum_{j=1}^{d} y_j, \ \boldsymbol{u} = \boldsymbol{y}/r, \quad \boldsymbol{y} \in \mathbb{R}_+^d. \tag{2.19}$$

In particular, $\boldsymbol{u}$ is the angular component of $\boldsymbol{y}$ that lies on the positive unit simplex

$$\mathbb{S}_+^{d-1} = \Big\{ u \in \mathbb{R}_+^d : \sum_{j=1}^{d} u_j = 1 \Big\}.$$

Next, we define $\omega = \omega^*$ to be a product of radial and angular measures

$$\omega(d\boldsymbol{y}) = \omega^*(dr, d\boldsymbol{u}) = \omega_r(dr) \times \omega_{\boldsymbol{u}}(d\boldsymbol{u}),$$

where ω_r is the standard Fréchet density

$$\omega_r(dr) := e^{-1/r} r^{-2} dr,$$

and $\omega_{\boldsymbol{u}}$ is a discrete measure over a finite subset $\mathcal{U} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\} \subset \mathbb{S}_+^{d-1}$. Specifically,

$$\omega_{\boldsymbol{u}}\left(d\boldsymbol{u}\right) := \sum_{\boldsymbol{u} \in \mathcal{U}} \delta_{\boldsymbol{u}}\left(d\boldsymbol{u}\right).$$

The specification of ω_r was chosen for analytical simplicity and as a matter of convention. The 1-Fréchet standardization (1.6) and the invariance property of max-linear combinations (1.8) imply that the CRPS criterion is roughly equivalent (up to a scaling factor) to the Cramér-von Mises distance along the radial direction. On the other hand, the choice of $\mathcal{U} \in \mathbb{S}_+^{d-1}$ can be rather arbitrary and the CRPS estimator remains consistent so long as $\mathcal{U}$ contains enough points for condition *(i)* of Theorem 2.4 (identifiability) to hold. We did find marginal improvement in CRPS based estimates with larger $|\mathcal{U}|$, especially if the dimension d is large. In the simulations that follow, we let $\mathcal{U}$ be a fixed random sample of size $m = 1000$ from the uniform distribution on $\mathbb{S}_+^{d-1}$.

With the weighting measure $\omega = \omega^*$ specified in such a way, the following lemma establishes a convenient closed form for the CRPS of max-stable models in terms of the tail dependence function $\sigma_{\boldsymbol{\theta}}$.

Lemma 2.8. *Suppose the measure ω in Definition 2.1 of the CRPS is specified as*

$$\omega^*(dr, d\boldsymbol{u}) = e^{-1/r} r^{-2} dr \sum_{\boldsymbol{u} \in \mathcal{U}} \delta_u(d\boldsymbol{u}), \tag{2.20}$$

where $\mathcal{U}$ is a finite subset of $\mathbb{S}_+^{d-1}$. Then the CRPS functional $\mathcal{E}_{\boldsymbol{\theta}}$ for a given max-stable model $F_{\boldsymbol{\theta}_0}(\boldsymbol{x}) = \exp(-\sigma_{\boldsymbol{\theta}_0}(\boldsymbol{x}))$, has the form $\mathcal{E}_{\boldsymbol{\theta}} = \mathcal{E}^*_{\boldsymbol{\theta}} + C$, where C is a constant that does not depend on $\boldsymbol{\theta}$, and

$$\begin{aligned}\mathcal{E}^*_{\boldsymbol{\theta}}(\boldsymbol{X}) &:= \sum_{\boldsymbol{u} \in \mathcal{U}} \int_0^\infty \left(e^{-\sigma_{\boldsymbol{\theta}}(\boldsymbol{u})/r} - \mathbf{1}_{\{X \le r\boldsymbol{u}\}}\right)^2 e^{-1/r} r^{-2} dr \\ &= \sum_{\boldsymbol{u} \in \mathcal{U}} \frac{1}{2\sigma_{\boldsymbol{\theta}}(\boldsymbol{u}) + 1} - \frac{2}{\sigma_{\boldsymbol{\theta}}(\boldsymbol{u}) + 1} \left(1 - \exp\left(\frac{\sigma_{\boldsymbol{\theta}}(\boldsymbol{u}) + 1}{M_{\boldsymbol{u}}}\right)\right), \end{aligned} \tag{2.21}$$

where

$$M_{\boldsymbol{u}} = \max_{j=1,\ldots,d} \{X_j / u_j\}. \tag{2.22}$$

See Section 2.6 for a proof.

In practice, given a set of independent observations $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ from the model $F_{\boldsymbol{\theta}_0}(\boldsymbol{x}) = \exp(-\sigma_{\boldsymbol{\theta}_0}(\boldsymbol{x}))$ we obtain the CRPS-based estimator of $\boldsymbol{\theta}_0$ as follows

CRPS estimation procedure

1. Construct the finite set $\mathcal{U} \subset \mathbb{S}^{d-1}$. This can be done heuristically or alternatively we found that large uniform random samples from the simplex $\mathbb{S}_+^{d-1}$ work well in a variety of circumstances.

2. Using numerical optimization, compute:

$$\hat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^n \mathcal{E}^*_{\boldsymbol{\theta}}(\boldsymbol{X}_i) \tag{2.23}$$

In Section 2.4, we illustrate this methodology over several concrete examples. The explicit construction of the set $\mathcal{U}$ is given in each example and the computation of

the tail dependence function $\sigma_{\boldsymbol{\theta}}$ when it is not available in closed form is discussed.

The following result provides readily computable expressions for the 'bread' and 'meat' matrices appearing in the asymptotic covariance of the CRPS estimators.

Corollary 2.9. *Suppose that the conditions on* $\boldsymbol{\theta} \mapsto F_{\boldsymbol{\theta}}(\boldsymbol{y}) = \exp(-\sigma_{\boldsymbol{\theta}}(\boldsymbol{y}))$ *in Proposition 2.6 hold with measure* ω^* *in* (2.20). *Define the random variable*

$$G_{\boldsymbol{u}} := (\sigma_{\boldsymbol{\theta}_0}(\boldsymbol{u}) + 1)^{-2} \int_0^{(\sigma_{\boldsymbol{\theta}_0}(\boldsymbol{u})+1)/M_{\boldsymbol{u}}} te^{-t} dt.$$

Then

$$\mathbf{H}_{\boldsymbol{\theta}_0} = 2 \sum_{\boldsymbol{u} \in \mathcal{U}} \frac{\dot{\sigma}_{\boldsymbol{\theta}_0}(\boldsymbol{u}) \left(\dot{\sigma}_{\boldsymbol{\theta}_0}(\boldsymbol{u})\right)^\top}{(2\sigma_{\boldsymbol{\theta}_0}(\boldsymbol{u}) + 1)^3}, \tag{2.24}$$

and

$$\mathbf{J}_{\boldsymbol{\theta}_0} = \sum_{\boldsymbol{u} \in \mathcal{U}} \sum_{w \in \mathcal{U}} c_{\boldsymbol{\theta}_0}(\boldsymbol{u}, \boldsymbol{w}) \dot{\sigma}_{\boldsymbol{\theta}_0}(\boldsymbol{u}) \left(\dot{\sigma}_{\boldsymbol{\theta}_0}(\boldsymbol{w})\right)^\top, \tag{2.25}$$

where

$$c_{\boldsymbol{\theta}_0}(\boldsymbol{u}, \boldsymbol{w}) = \mathrm{Cov}\left(G_{\boldsymbol{u}}, G_{\boldsymbol{w}}\right).$$

Remark 2.10. $M_{\boldsymbol{u}}$ and $M_{\boldsymbol{w}}$ are dependent (and so are $G_{\boldsymbol{u}}$ and $G_{\boldsymbol{w}}$) since in view of (2.22) they are defined as max-linear combinations of the same vector $\boldsymbol{X}$. The coefficient $c_{\boldsymbol{\theta}_0}(\boldsymbol{u}, \boldsymbol{w})$ can be conveniently computed using Monte Carlo methods by simulating a large number of independent copies of X under the $F_{\boldsymbol{\theta}_0}$ model. In practice the resulting asymptotic covariance matrix estimates yield confidence intervals with close to nominal coverage (see Tables 2.2 and 2.3).

2.4 Simulation

In this section, we conduct simulation studies for CRPS M-estimation under 3 different max-stable models. The first example provides a comparison of CRPS M-estimation to the MLE in the case of the multivariate logistic model. The second

example illustrates inference for a random field model applicable in spatial extremes. Here CRPS M-estimation is compared with the popular pairwise maximum composite likelihood estimator of Padoan et al. (2010). The final example illustrates CRPS M-estimation for max-linear models where pairwise likelihoods are not available and compares it to the alternative M-estimator in Einmahl et al. (2012).

2.4.1 Example: multivariate logistic model

The multivariate logistic is a special case that allows comparison between our CRPS based estimator and the MLE, since the full joint likelihood is available in this simple model. Hence, we can estimate the relative efficiency of the CRPS estimator in this idealized case. To this end, let $\boldsymbol{\theta} = (\lambda, \alpha) \in \Theta := (0, \infty) \times (0, 1)$ and recall

$$\sigma_{\boldsymbol{\theta}}(\boldsymbol{x}) = \lambda \Big(\sum_{i=1}^{d} x_i^{-1/\alpha} \Big)^{\alpha},$$

is the tail dependence function of a multivariate logistic max-stable model. We estimate the parameters for the model when $d = 5$ and $\boldsymbol{\theta}_0 = (5, 0.7)$, using sample sizes $n = 100$ and $n = 1000$ with 500 replications each. Realizations were generated using the `R` package `evd` (Stephenson, 2002). For each realization $\boldsymbol{X}_i, i = 1, \ldots, n$ we construct the max-linear combinations $M_{\boldsymbol{u}}^{(i)}$ using a (fixed) uniform sample $\mathcal{U} \subset \mathbb{S}^{d-1}$ where $|\mathcal{U}| = 1000$. Numerical optimization of the CRPS criterion in (2.23) was carried out using `R`'s `optim` routine with an arbitrary starting point in the interior of Θ. Results for both the CRPS estimators and the MLE are shown in Table 2.2.

Observe that we have essentially unbiased estimators. The asymptotic confidence intervals based on (2.14) were computed using the expressions in Corollary 2.9 and have close to nominal coverages even for moderate sample size $n = 100$. As expected, the CRPS is less efficient than the MLE however, the results in Table 2.2 provide evidence that suggest the CRPS is a good alternative when the MLE is not available.

Table 2.2: Logistic model simulation results using 500 replications of sample size $n = 100$ and $n = 1000$. Reported are bias and root mean squared error (RMSE) for CRPS and MLE estimates. Coverages are based on plug-in estimates of 95% asymptotic confidence intervals. In the case of the CRPS estimates, confidence intervals are based on (2.14) and computed using the expressions from Corollary 2.9.

		Bias		RMSE		95% Coverage	
		CRPS	MLE	CRPS	MLE	CRPS	MLE
$n = 100$	$\lambda = 5.0$	0.0200	0.0406	0.3706	0.3401	0.958	0.940
	$\alpha = 0.7$	0.0053	0.0025	0.0481	0.0259	0.962	0.948
$n = 1000$	$\lambda = 5.0$	0.0010	0.0010	0.1230	0.1060	0.940	0.938
	$\alpha = 0.7$	0.0001	0.0003	0.0144	0.0082	0.948	0.940

2.4.2 Example: Schlather model

We now provide an example that is applicable in the spatial setting. Let $\{W(s)\}_{s\in S}$ be a Gaussian process on $S \subset \mathbb{R}^2$ with standard normal margins and let $\rho_{\boldsymbol{\theta}}(t, s)$ be its associated correlation function parameterized by $\boldsymbol{\theta}$. Suppose we observe the process at a finite set of 'observation' locations $\{s_1, \ldots, s_d\} = D$. Define

$$\sigma_{\boldsymbol{\theta}}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{\theta}} \max_{s\in D} \left\{ (\sqrt{2\pi} W(s) \vee 0)/x_s \right\}, \tag{2.26}$$

Where $\mathbb{E}_{\boldsymbol{\theta}}$ denotes expectation with respect to the density of $\{W(s)\}_{s\in D}$ specified by the parameter $\boldsymbol{\theta}$. $\sigma_{\boldsymbol{\theta}}(\boldsymbol{x})$ is the tail dependence function (on the set D) of a Schlather max-stable model with standard 1-Fréchet marginals. In this case $\sigma_{\boldsymbol{\theta}}(\boldsymbol{x})$ is not available in closed form, instead we use a Monte Carlo approximation to the expectation in (2.26) using a large sample $W_i(s)$, $i = 1, \ldots, K = 10^5$ under $\boldsymbol{\theta}$. For this simulation we assume a *stable* correlation function, i.e.

$$\rho_{\boldsymbol{\theta}}(t, s) = \exp\left[-\left(\|t - s\| / \lambda \right)^{\alpha} \right], \ \boldsymbol{\theta} = (\lambda, \alpha) \in \Theta = (0, \infty) \times (0, 2].$$

Table 2.3: CRPS and MCLE estimates for Schlather model. Reported are bias and root mean squared error (RMSE) of 500 replications using sample size $n \in \{100, 1000, 5000\}$. CRPS based confidence intervals for $\boldsymbol{\theta}_0 = (100, 1)$ were calculated using plug-in estimates for the expressions in Corollary 2.9 and resulting 95% coverages are reported. Coverages for MCLE estimates are based on the Godambe information (see Padoan et al., 2010).

		Bias		RMSE		95% Coverage	
		CRPS	MCLE	CRPS	MCLE	CRPS	MCLE
$n = 100$	$\lambda = 100$	31.89	2.01	156.86	16.57	0.988	0.948
	$\alpha = 1$	0.058	0.005	0.565	0.171	0.860	0.922
$n = 1000$	$\lambda = 100$	-1.2264	0.1704	25.2951	4.8049	0.970	0.950
	$\alpha = 1$	0.0246	0.0000	0.2511	0.0549	0.908	0.944
$n = 5000$	$\lambda = 100$	-0.6179	0.1059	9.3670	2.0762	0.960	0.956
	$\alpha = 1$	0.0086	0.0001	0.1159	0.0243	0.932	0.940

The top row of Figure 2.1 shows realizations from this Schlather model under two different parameter settings. For our study we set $\boldsymbol{\theta}_0 = (100, 1)$ and simulated 500 replications at $d = 30$ uniformly sampled locations over a 500×500 grid. This corresponds to the top left panel in Figure 2.1.

Realizations were generated using the `R` package `SpatialExtremes` (Ribatet, 2011). For each realization $\boldsymbol{X}_i, i = 1, \ldots, n$ we construct the max-linear combinations $M_{\boldsymbol{u}}^{(i)}$ using a random uniform sample $\mathcal{U} \subset \mathbb{S}^{d-1}$, where $|\mathcal{U}| = 1000$. For sample sizes $n \in \{100, 1000, 5000\}$, we numerically optimize the CRPS criterion (2.23) using `R`'s `optim` routine with multiple starting points in the interior of Θ. Simulation results in Table 2.3 show that with the exception of α at small sample size, CRPS estimates are essentially unbiased and confidence intervals display close to nominal coverage. For comparison we also provide pairwise MCLE estimates fitted using the `SpatialExtremes` package. For information on pairwise MCLE see Padoan et al. (2010).

In this case pairwise MCLE outperforms the CRPS. This is an instance of a

tradeoff between universality and efficiency. While our CRPS estimators are generally applicable to a wide variety of max-stable models, likelihood based methods are more efficient when available. The following example shows a case where the MCLE is unavailable.

2.4.3 Example: max-linear model

To illustrate our CRPS based estimation with spectrally discrete max-stable models we consider the four dimensional two-factor model used in Einmahl et al. (2012). Let $A = (a_{jk})_{4\times 2}$ be a matrix with non-negative entries such that $a_{j1} + a_{j2} = 1,\ j = 1, \ldots, 4$. Define

$$X_j = a_{j1}Z_1 \vee a_{j2}Z_2,\ j = 1, \ldots, 4$$

where Z_1 and Z_2 are independent standard Fréchet. The vector $\boldsymbol{X} = (X_j)_{j=1}^4$ is max-stable with standard Fréchet margins and tail dependence function

$$\sigma(\boldsymbol{x}) = \sum_{k=1}^{2} \max_{j=1,\ldots,4} a_{jk}/x_j.$$

The row sum condition on A implies four parameters to estimate. To avoid identifiability issues resulting from permuting Z_1 and Z_2, we define the parameters as the column of A with largest sum

$$\theta_{0j} := a_{jk^*},\ k^* = \arg\max_{k\in\{1,2\}} \sum_{i=1}^{d} a_{ik},\ j = 1, \ldots, 4.$$

In the case of a tie among the column sums, there is no ambiguity as either column specifies the same parameter.

We simulated 500 replications from the model with $\boldsymbol{\theta}_0 = (0.2, 0.5, 0.7, 0.9)^\top$ and using a sample size $n = 5000$. For each realization $\boldsymbol{X}_i, i = 1, \ldots, n$ we construct the max-linear combinations $M_{\boldsymbol{u}}^{(i)}$ using a random uniform sample $\mathcal{U} \subset \mathbb{S}^{d-1}$, where

Table 2.4: Simulation results of the four dimensional two factor max-linear model using 500 replications of sample size $n = 5000$. We compare the CRPS estimator with the M-estimator (M-est) of Einmahl et al. (2012). Their estimator depends on a threshold parameter $\kappa \in (0, n)$, thus reported values are favorable ranges based off of the graphs in Figure 5 of Einmahl et al. (2012) which plot bias and root mean squared error (RMSE) over a wide range of $\kappa \in [40, 1000]$.

	Bias		RMSE	
	CRPS	M-est	CRPS	M-est
$\theta_{01} = 0.2$	0.0005	$(0.000, 0.040)$	0.0176	$(0.020, 0.050)$
$\theta_{02} = 0.5$	0.0004	$(-0.005, 0.002)$	0.0080	$(0.010, 0.050)$
$\theta_{03} = 0.7$	0.0012	$(-0.005, 0.000)$	0.0131	$(0.010, 0.045)$
$\theta_{04} = 0.9$	0.0011	$(-0.035, 0.000)$	0.0182	$(0.020, 0.040)$

$|\mathcal{U}| = 1000$. We numerically optimize the CRPS criterion (2.23) using `R`'s `optim` routine. Results are shown in Table 2.4. For comparison we report values from the identical simulation conducted in Einmahl et al. (2012). In this case the CRPS estimation has root mean squared error (RMSE) that is almost uniformly lower than the competing M-estimator. Indeed, this could be due to the fact that the M-estimator of Einmahl et al. (2012) makes a weaker assumption in that it does not assume exact max-stable distributions rather simply distributions in the max domain of attraction of a suitable max-stable model.

2.5 Discussion

In this Chapter we have developed a general inferential framework for max-stable models based on the continuous ranked probability score (CRPS). It is shown that under mild regularity conditions, CRPS M-estimators are consistent and asymptotically normal. Simulation studies across popular max-stable models yield essentially unbiased estimators with close to nominal coverage. Our simulation results indicate

that the CRPS has lower asymptotic efficiency than likelihood based methods but remains an attractive alternative when likelihoods or composite likelihoods are not available.

From a computational aspect, if the parameter space Θ is small, the CRPS based method can easily handle models with dimension d in the 100's. In practice, however, in very high-dimensional settings, the CRPS-based statistics may lack the asymptotic efficiency required. Therefore, the method and especially the choice of the weighting measure ω needs to be applied with care. The study of the optimal choice of the measure ω in general (or the set $\mathcal{U}$ in particular) is rather challenging. For example, one could try to optimize a matrix norm of the asymptotic covariance matrix appearing in Theorem 2.5. This would involve solving an optimization problem on the infinite dimensional space of possible measures. This is an interesting optimization problem that is studied under a different context in Chapter 5

2.6 Proofs

Proof of Theorem 2.4. Observe that the estimator $\hat{\boldsymbol{\theta}}_n$ in Definition 2.1 trivially satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{E}_{\hat{\boldsymbol{\theta}}_n}\left(\boldsymbol{X}_i\right) \leq \frac{1}{n}\sum_{i=1}^{n}\mathcal{E}_{\boldsymbol{\theta}_0}\left(\boldsymbol{X}_i\right) - o_p\left(1\right).$$

Therefore, by Thm. 5.7 of van der Vaart, 1998, the desired consistency follows if

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{n}\sum_{i=1}^{n}\mathcal{E}_{\boldsymbol{\theta}}\left(\boldsymbol{X}_i\right) - \mathbb{E}\mathcal{E}_{\boldsymbol{\theta}}\left(\boldsymbol{X}\right)\right| \xrightarrow{p} 0 \tag{2.27}$$

and

$$\sup_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\geq\varepsilon,\ \boldsymbol{\theta}\in\Theta}\mathbb{E}\mathcal{E}_{\boldsymbol{\theta}}\left(\boldsymbol{X}\right) > \mathbb{E}\mathcal{E}_{\boldsymbol{\theta}_0}\left(\boldsymbol{X}\right), \quad \text{for all } \varepsilon > 0. \tag{2.28}$$

We will first show (2.28). By Fubini's Theorem, we have

$$\mathbb{E}\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X}) = \int_{\mathbb{R}^d} \left(F_{\boldsymbol{\theta}}(\boldsymbol{y}) - F_{\boldsymbol{\theta}_0}(\boldsymbol{y})\right)^2 \omega(d\boldsymbol{y}) + \int_{\mathbb{R}^d} F_{\boldsymbol{\theta}_0}(\boldsymbol{y})\left(1 - F_{\boldsymbol{\theta}_0}(\boldsymbol{y})\right)\omega(d\boldsymbol{y}) \geq \int_{\mathbb{R}^d} F_{\boldsymbol{\theta}_0}(\boldsymbol{y})\left(1 - F_{\boldsymbol{\theta}_0}(\boldsymbol{y})\right)\omega(d\boldsymbol{y}) = \mathbb{E}\mathcal{E}_{\boldsymbol{\theta}_0}(\boldsymbol{X}). \quad (2.29)$$

This implies (2.28) because the continuity condition *(iii)* and the compactness of Θ gaurantee the supremum therein is attained for some $\boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0$.

We now show (2.27). Let $F_n(\boldsymbol{x}) = n^{-1}\sum_{i=1}^n \mathbf{1}\{\boldsymbol{X}_i \leq \boldsymbol{x}\}$ and $\overline{F} = 1 - F$. Note that

$$\sup_{\boldsymbol{\theta}\in\Theta}\left|\frac{1}{n}\sum_{i=1}^n \mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X}_i) - \mathbb{E}\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X})\right| = \sup_{\boldsymbol{\theta}\in\Theta}\left|\int_{\mathbb{R}^d}\left(1 - 2F_{\boldsymbol{\theta}}(\boldsymbol{x})\right)\left(F_n(\boldsymbol{x}) - F_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right)\omega(d\boldsymbol{x})\right| \leq \int_{\mathbb{R}^d}\left|F_n(\boldsymbol{x}) - F_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right|\omega(d\boldsymbol{x}). \quad (2.30)$$

Fix $\epsilon > 0$. Markov's inequality and another application of Fubini gives

$$\mathbb{P}\left\{\int_{\mathbb{R}^d}\left|F_n(\boldsymbol{x}) - F_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right|\omega(d\boldsymbol{x}) > \epsilon\right\} \leq \frac{1}{\epsilon}\int_{\mathbb{R}^d}\mathbb{E}\left|\overline{F}_n(\boldsymbol{x}) - \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right|\omega(d\boldsymbol{x}). \quad (2.31)$$

Next, using the identity $|a-b| = a+b-2a\wedge b$ and the fact that $\mathbb{E}\overline{F}_n(\boldsymbol{x}) = \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})$, we have that the RHS of (2.31) equals

$$\frac{1}{\epsilon}\int_{\mathbb{R}^d}\mathbb{E}\left\{\overline{F}_n(\boldsymbol{x}) + \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x}) - 2\overline{F}_n(\boldsymbol{x})\wedge\overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right\}\omega(d\boldsymbol{x}) = \frac{2}{\epsilon}\left\{\int_{\mathbb{R}^d}\overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\omega(d\boldsymbol{x}) - \int_{\mathbb{R}^d}\mathbb{E}\left[\overline{F}_n(\boldsymbol{x})\wedge\overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right]\omega(d\boldsymbol{x})\right\}. \quad (2.32)$$

The strong law of large numbers implies that $\overline{F}_n(\boldsymbol{x}) \wedge \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})$ converges almost surely to $\overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x}) \wedge \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x}) \equiv \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})$. Hence, by applying the Lebesgue dominated convergence theorem, we have

$$\lim_{n\to\infty} \mathbb{E}\left[\overline{F}_n(\boldsymbol{x}) \wedge \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right] = \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x}), \quad \text{for all } x \in \mathbb{R}^d. \tag{2.33}$$

Note that $\mathbb{E}\left[\overline{F}_n(\boldsymbol{x}) \wedge \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right] \leq \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})$, and by condition *(ii)*, $\int_{\mathbb{R}^d} \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\, \omega(d\boldsymbol{x}) < \infty$. Thus, by a second application of DCT

$$\begin{aligned}
&\lim_{n\to\infty} \int_{\mathbb{R}^d} \mathbb{E}\left[\overline{F}_n(\boldsymbol{x}) \wedge \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right] \omega(d\boldsymbol{x}) \\
&\qquad = \int_{\mathbb{R}^d} \lim_{n\to\infty} \mathbb{E}\left[\overline{F}_n(\boldsymbol{x}) \wedge \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\right] \omega(d\boldsymbol{x}) \overset{(2.33)}{=} \int_{\mathbb{R}^d} \overline{F}_{\boldsymbol{\theta}_0}(\boldsymbol{x})\, \omega(d\boldsymbol{x}).
\end{aligned} \tag{2.34}$$

This, by (2.32) implies that the right-hand side of (2.31) vanishes as $n \to \infty$, which in view of (2.30) yields the desired convergence in probability (2.27) and the proof is complete. □

Proof of Theorem 2.5. Since the CRPS estimator $\hat{\boldsymbol{\theta}}_n$ minimizes the CRPS distance, we trivially have $n^{-1}\sum_{i=1}^n \mathcal{E}_{\hat{\boldsymbol{\theta}}_n}(\boldsymbol{X}_i) \leq n^{-1}\sum_{i=1}^n \mathcal{E}_{\boldsymbol{\theta}_0}(\boldsymbol{X}_i) - o_p(n^{-1})$. Thus, by Thm. 5.23 of van der Vaart, 1998 the asymptotic normality in (2.14) follows, provided conditions (i)-(iii) hold. □

Proof of Proposition 2.6. By a standard argument using the Lebesgue DCT, condition *(iii)* of this proposition ensures that integration and differentiation can be interchanged in all that follows. We proceed by establishing *(i)*-*(iii)* of Theorem 2.5.

(i) By the differentiability of $\boldsymbol{\theta} \mapsto F_{\boldsymbol{\theta}}$ for all $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0)$ the function $\boldsymbol{\theta} \mapsto \mathcal{E}_{\boldsymbol{\theta}}$ is

differentiable at $\boldsymbol{\theta}_0$ since exchanging integration and differentiation allows

$$\begin{aligned}\dot{\mathcal{E}}_{\boldsymbol{\theta}_0} &= \frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbb{R}^d} \left(F_{\boldsymbol{\theta}}(\boldsymbol{y}) - \mathbf{1}\{\boldsymbol{x} \leq \boldsymbol{y}\}\right)^2 \omega(d\boldsymbol{y})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= 2\int_{\mathbb{R}^d} \left(F_{\boldsymbol{\theta}_0}(\boldsymbol{y}) - \mathbf{1}\{\boldsymbol{x} \leq \boldsymbol{y}\}\right) \dot{F}_{\boldsymbol{\theta}_0}(\boldsymbol{y})\, \omega(d\boldsymbol{y}).\end{aligned}$$

(ii) Observe that $|\mathcal{E}_{\boldsymbol{\theta}_1}(\boldsymbol{x}) - \mathcal{E}_{\boldsymbol{\theta}_2}(\boldsymbol{x})|$ equals

$$\begin{aligned}&\left|\int_{\mathbb{R}^d} \left\{\left(F_{\boldsymbol{\theta}_1}(\boldsymbol{y}) - \mathbf{1}\{\boldsymbol{x} \leq \boldsymbol{y}\}\right)^2 - \left(F_{\boldsymbol{\theta}_2}(\boldsymbol{y}) - \mathbf{1}\{\boldsymbol{x} \leq \boldsymbol{y}\}\right)^2\right\} \omega(d\boldsymbol{y})\right| \\ &\quad = \left|\int_{\mathbb{R}^d} \left\{\left[\left(F_{\boldsymbol{\theta}_1}(\boldsymbol{y}) + F_{\boldsymbol{\theta}_2}(\boldsymbol{y})\right) - 2\mathbf{1}\{\boldsymbol{x} \leq \boldsymbol{y}\}\right]\left(F_{\boldsymbol{\theta}_1}(\boldsymbol{y}) - F_{\boldsymbol{\theta}_2}(\boldsymbol{y})\right)\right\} \omega(d\boldsymbol{y})\right| \\ &\quad \leq 2\int_{\mathbb{R}^d} \left|F_{\boldsymbol{\theta}_1}(\boldsymbol{y}) - F_{\boldsymbol{\theta}_2}(\boldsymbol{y})\right| \omega(d\boldsymbol{y})\end{aligned}$$

where the last relation follows from the triangle inequality and the fact that

$$\left|F_{\boldsymbol{\theta}}(\boldsymbol{y}) - \mathbf{1}\{\boldsymbol{x} \leq \boldsymbol{y}\}\right| \leq \max\left\{F_{\boldsymbol{\theta}}(\boldsymbol{y}), 1 - F_{\boldsymbol{\theta}}(\boldsymbol{y})\right\} \leq 1.$$

Then, by the mean value theorem and the Cauchy-Schwartz inequality

$$\begin{aligned}\int_{\mathbb{R}^d} \left|F_{\boldsymbol{\theta}_1}(\boldsymbol{y}) - F_{\boldsymbol{\theta}_2}(\boldsymbol{y})\right| \omega(d\boldsymbol{y}) &\leq \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\| \int_{\mathbb{R}^d} \sup_{\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0)} \left\|\dot{F}_{\boldsymbol{\theta}}(\boldsymbol{y})\right\| \omega(d\boldsymbol{y}) \\ &\equiv L\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|\end{aligned} \tag{2.35}$$

where $L := \int_{\mathbb{R}^d} \sup_{\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0)} \left\|\dot{F}_{\boldsymbol{\theta}}(\boldsymbol{y})\right\| \omega(d\boldsymbol{y})$. By assumption *(ii)* of this proposition, L is finite. Hence *(ii)* of Theorem 2.5 holds where $L(\boldsymbol{X}) \equiv L$ is constant (and therefore trivially $\mathbb{E}\left(L(\boldsymbol{X})^2\right) < \infty$).

(iii) Existence of a second order Taylor expansion for $\boldsymbol{\theta} \mapsto \mathbb{E}\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X})$ follows from the twice continuous differentiability of $\boldsymbol{\theta} \mapsto F_{\boldsymbol{\theta}}$ for all $\boldsymbol{\theta} \in B(\boldsymbol{\theta}_0)$ by

$$\begin{aligned}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\mathbb{E}\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X}) &\overset{(2.29)}{=} \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\int_{\mathbb{R}^d}\left(F_{\boldsymbol{\theta}}(\boldsymbol{y})-F_{\boldsymbol{\theta}_0}(\boldsymbol{y})\right)^2\omega(d\boldsymbol{y})\\ &= \int_{\mathbb{R}^d}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\left(F_{\boldsymbol{\theta}}(\boldsymbol{y})-F_{\boldsymbol{\theta}_0}(\boldsymbol{y})\right)^2\omega(d\boldsymbol{y}).\end{aligned}$$

The above display implies that

$$\begin{aligned}\mathbf{H}_{\boldsymbol{\theta}_0} = \int_{\mathbb{R}^d}\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\left(F_{\boldsymbol{\theta}}(\boldsymbol{y})-F_{\boldsymbol{\theta}_0}(\boldsymbol{y})\right)^2\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\omega(d\boldsymbol{y})\\ = 2\int_{\mathbb{R}^d}\dot{F}_{\boldsymbol{\theta}_0}(\boldsymbol{y})\dot{F}_{\boldsymbol{\theta}_0}(\boldsymbol{y})^\top\omega(d\boldsymbol{y}) = (2.17)\end{aligned}$$

where non-singularity of $\mathbf{H}_{\boldsymbol{\theta}_0}$ follows from *(ii)* because for all $a\in\mathbb{R}^p$ with $\|a\|>0$

$$a^\top\mathbf{H}_{\boldsymbol{\theta}_0}a = 2\int_{\mathbb{R}^d}\left[a^\top\dot{F}(\boldsymbol{y})\right]^2\omega(d\boldsymbol{y})>0.$$

Finally, we derive $\mathbf{J}_{\boldsymbol{\theta}_0}$ by considering its ijth entry. Let ∂_i denote $\partial/\partial\boldsymbol{\theta}_i$.

$$\begin{aligned}(\mathbf{J}_{\boldsymbol{\theta}_0})_{ij} &= \mathbb{E}\left[\partial_i\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X})\,\partial_j\mathcal{E}_{\boldsymbol{\theta}}(\boldsymbol{X})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right]\\ &= \mathbb{E}\left\{\int_{\mathbb{R}^d}2\left(F_{\boldsymbol{\theta}}(\boldsymbol{y}_1)-\mathbf{1}_{\{X\le\boldsymbol{y}_1\}}\right)\partial_iF_{\boldsymbol{\theta}}(\boldsymbol{y}_1)\,\omega(d\boldsymbol{y}_1)\right.\\ &\qquad\left.\times\int_{\mathbb{R}^d}2\left(F_{\boldsymbol{\theta}}(\boldsymbol{y}_2)-\mathbf{1}_{\{X\le\boldsymbol{y}_2\}}\right)\partial_jF_{\boldsymbol{\theta}}(\boldsymbol{y}_2)\,\omega(d\boldsymbol{y}_2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\}\\ &= 4\mathbb{E}\left\{\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}b_{\boldsymbol{\theta}}(X,\boldsymbol{y}_1,\boldsymbol{y}_2)\,\partial_iF_{\boldsymbol{\theta}}(\boldsymbol{y}_1)\,\partial_jF_{\boldsymbol{\theta}}(\boldsymbol{y}_2)\,\omega(d\boldsymbol{y}_1)\,\omega(d\boldsymbol{y}_2)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right\}\end{aligned}$$

where $b_{\boldsymbol{\theta}}(X,\boldsymbol{y}_1,\boldsymbol{y}_2)=\left(\mathbf{1}_{\{X\le\boldsymbol{y}_1\}}-F_{\boldsymbol{\theta}}(\boldsymbol{y}_1)\right)\left(\mathbf{1}_{\{\boldsymbol{X}\le\boldsymbol{y}_2\}}-F_{\boldsymbol{\theta}}(\boldsymbol{y}_2)\right)$. Expanding the integrand and applying Fubini gives

$$(\mathbf{J}_{\boldsymbol{\theta}_0})_{ij} = 4\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\beta_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1,\boldsymbol{y}_2)\,\partial_iF_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1)\,\partial_jF_{\boldsymbol{\theta}_0}(\boldsymbol{y}_2)\,\omega(d\boldsymbol{y}_1)\,\omega(d\boldsymbol{y}_2)$$

where $\beta_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1,\boldsymbol{y}_2)=\mathbb{E}b_{\boldsymbol{\theta}_0}(\boldsymbol{X},\boldsymbol{y}_1,\boldsymbol{y}_2)=F_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1\wedge\boldsymbol{y}_2)-F_{\boldsymbol{\theta}_0}(\boldsymbol{y}_1)F_{\boldsymbol{\theta}_0}(\boldsymbol{y}_2)$, which is

exactly the ijth element of (2.18), as desired. □

Proof of Lemma 2.8. Recall $M_{\boldsymbol{u}} := \max_{j=1,\dots,d}\{X_j/u_j\}$. Fix $u \in \mathcal{U}$. Note that $e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{u})/r} - \mathbf{1}_{\{\boldsymbol{X}\leq r\boldsymbol{u}\}}$ is bounded and hence the integral in (2.21) is finite. Observing that $\{\boldsymbol{X} \leq r\boldsymbol{u}\} = \{M_u \leq r\}$ and making the change of variables $t = 1/r$, we have that

$$\begin{aligned}\int_0^\infty \left(e^{-V_{\boldsymbol{\theta}}(\boldsymbol{u})/r} - \mathbf{1}_{\{\boldsymbol{X}\leq r\boldsymbol{u}\}}\right)^2 e^{-1/r} r^{-2} dr &= \int_0^\infty \left(e^{-V_{\boldsymbol{\theta}}(\boldsymbol{u})t} - \mathbf{1}_{\{t\leq M_{\boldsymbol{u}}^{-1}\}}\right)^2 e^{-t} dt \\ &= \int_0^{M_{\boldsymbol{u}}^{-1}} \left(e^{-2V_{\boldsymbol{\theta}}(\boldsymbol{u})t} - 2e^{-V_{\boldsymbol{\theta}}(\boldsymbol{u})t} + 1\right) e^{-t} dt + \int_{M_{\boldsymbol{u}}^{-1}}^\infty e^{-2V_{\boldsymbol{\theta}}(\boldsymbol{u})t} e^{-t} dt \\ &= \int_0^\infty e^{-(2V_{\boldsymbol{\theta}}(\boldsymbol{u})+1)t} dt - 2\int_0^{M_{\boldsymbol{u}}^{-1}} e^{-(V_{\boldsymbol{\theta}}(\boldsymbol{u})+1)t} dt + \int_0^{M_{\boldsymbol{u}}^{-1}} e^{-t} dt.\end{aligned}$$

All three integrals above have trivial closed form expressions which yield

$$\frac{1}{2V_{\boldsymbol{\theta}}(\boldsymbol{u})+1} - \frac{2}{V_{\boldsymbol{\theta}}(\boldsymbol{u})+1}\left(1 - \exp\left(-\frac{V_{\boldsymbol{\theta}}(\boldsymbol{u})+1}{M_{\boldsymbol{u}}}\right)\right) + 1 - \exp\left(-\frac{1}{M_{\boldsymbol{u}}}\right).$$

Finally, terms that do not depend on $\boldsymbol{\theta}$ can be ignored giving the desired expression in (2.21). □

Proof of Corollary 2.9. We first establish expression (2.24) for $\mathbf{H}_{\boldsymbol{\theta}_0}$. Substituting the measure ω^* from (2.20) into the expression (2.17) for $\mathbf{H}_{\boldsymbol{\theta}_0}$, we have

$$\mathbf{H}_{\boldsymbol{\theta}_0} = \sum_{u\in\mathcal{U}} \int_0^\infty \dot{F}_{\boldsymbol{\theta}_0}(r\boldsymbol{u}) \left(\dot{F}_{\boldsymbol{\theta}_0}(r\boldsymbol{u})\right)^\top e^{-1/r} r^{-2} dr. \tag{2.36}$$

Under a max-stable model with 1-Fréchet margins, the homogeneity property of the tail dependence function gives

$$\dot{F}_{\boldsymbol{\theta}_0}(r\boldsymbol{u}) = \frac{\partial}{\partial\boldsymbol{\theta}} e^{-V_{\boldsymbol{\theta}}(r\boldsymbol{u})}\bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = -e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{u})/r} \frac{\dot{V}_{\boldsymbol{\theta}_0}(\boldsymbol{u})}{r}. \tag{2.37}$$

Substituting (2.37) into (2.36) yields

$$\mathbf{H}_{\boldsymbol{\theta}_0} = \sum_{u \in \mathcal{U}} \dot{V}_{\boldsymbol{\theta}_0}(\boldsymbol{u}) \left(\dot{V}_{\boldsymbol{\theta}_0}(\boldsymbol{u}) \right)^\top \int_0^\infty e^{-(2V_{\boldsymbol{\theta}}(\boldsymbol{u})+1)/r} r^{-4} dr. \tag{2.38}$$

Finally, by substituting $t = 1/r$, the integral in (2.38) has closed form $2/(2V_{\boldsymbol{\theta}}(\boldsymbol{u})+1)^3$ which implies

$$\mathbf{H}_{\boldsymbol{\theta}_0} = 2 \sum_{u \in \mathcal{U}} \frac{\dot{V}_{\boldsymbol{\theta}_0}(\boldsymbol{u}) \left(\dot{V}_{\boldsymbol{\theta}_0}(\boldsymbol{u}) \right)^\top}{(2V_{\boldsymbol{\theta}}(\boldsymbol{u}) + 1)^3}$$

This concludes the proof for $\mathbf{H}_{\boldsymbol{\theta}_0}$.

We now establish (2.25) for $\mathbf{J}_{\boldsymbol{\theta}_0}$. Substituting the measure ω^* from (2.20) into the expression (2.18) for $\mathbf{J}_{\boldsymbol{\theta}_0}$, we have

$$\mathbf{J}_{\boldsymbol{\theta}_0} = \sum_{u \in \mathcal{U}} \sum_{w \in \mathcal{U}} \int_0^\infty \int_0^\infty \beta_{\boldsymbol{\theta}_0}(r\boldsymbol{u}, s\boldsymbol{w}) \dot{F}_{\boldsymbol{\theta}_0}(r\boldsymbol{u}) \left(\dot{F}_{\boldsymbol{\theta}_0}(s\boldsymbol{w}) \right)^\top \omega_r(dr) \omega_r(ds) \tag{2.39}$$

where $\beta_{\boldsymbol{\theta}_0}(r\boldsymbol{u}, s\boldsymbol{w}) = F_{\boldsymbol{\theta}_0}(r\boldsymbol{u} \wedge s\boldsymbol{w}) - F_{\boldsymbol{\theta}_0}(r\boldsymbol{u}) F_{\boldsymbol{\theta}_0}(s\boldsymbol{w})$ and $\omega_r(dr) = e^{-1/r} r^{-2} dr$. In view of (2.37) this gives

$$\mathbf{J}_{\boldsymbol{\theta}_0} = \sum_{u \in \mathcal{U}} \sum_{w \in \mathcal{U}} \dot{V}_{\boldsymbol{\theta}_0}(\boldsymbol{u}) \left(\dot{V}_{\boldsymbol{\theta}_0}(\boldsymbol{w}) \right)^\top \int_0^\infty \int_0^\infty \frac{1}{rs} \beta_{\boldsymbol{\theta}_0}(r\boldsymbol{u}, s\boldsymbol{w}) F_{\boldsymbol{\theta}_0}(r\boldsymbol{u}) F_{\boldsymbol{\theta}_0}(s\boldsymbol{w}) \omega_r(dr) \omega_r(ds). \tag{2.40}$$

Now recall $G_{\boldsymbol{u}} = (V_{\boldsymbol{\theta}_0}(\boldsymbol{u})+1)^{-2} \int_0^{(V_{\boldsymbol{\theta}_0}(\boldsymbol{u})+1)/M_{\boldsymbol{u}}} t e^{-t} dt$. To complete the proof, we must show

$$\mathrm{Cov}(G_{\boldsymbol{u}}, G_{\boldsymbol{w}}) = \int_0^\infty \int_0^\infty \frac{1}{rs} \beta_{\boldsymbol{\theta}_0}(r\boldsymbol{u}, s\boldsymbol{w}) F_{\boldsymbol{\theta}_0}(r\boldsymbol{u}) F_{\boldsymbol{\theta}_0}(s\boldsymbol{w}) \omega_r(dr) \omega_r(ds).$$

This is equivalent to showing

$$\mathbb{E} G_{\boldsymbol{u}} = \int_0^\infty \frac{1}{r} \left(F_{\boldsymbol{\theta}_0}(r\boldsymbol{u}) \right)^2 \omega_r(dr), \tag{2.41}$$

and

$$\mathbb{E}\{G_{\boldsymbol{u}}G_{\boldsymbol{w}}\} = \int_0^\infty \int_0^\infty \frac{1}{rs} F_{\boldsymbol{\theta}_0}(r\boldsymbol{u} \wedge s\boldsymbol{w}) F_{\boldsymbol{\theta}_0}(r\boldsymbol{u}) F_{\boldsymbol{\theta}_0}(s\boldsymbol{w}) \omega_r(dr)\omega_r(ds). \tag{2.42}$$

We first establish (2.41). We begin with

$$\mathbb{E}G_{\boldsymbol{u}} = (V_{\boldsymbol{\theta}_0}(\boldsymbol{u}) + 1)^{-2}\mathbb{E}\int_0^{(V_{\boldsymbol{\theta}_0}(\boldsymbol{u})+1)/M_{\boldsymbol{u}}} te^{-t}dt.$$

Substituting $t = (V_{\boldsymbol{\theta}_0}(\boldsymbol{u}) + 1)/r$ yields

$$\mathbb{E}G_{\boldsymbol{u}} = \mathbb{E}\int_{M_{\boldsymbol{u}}}^\infty e^{-(V_{\boldsymbol{\theta}_0}(\boldsymbol{u})+1)/r} r^{-3}dr = \mathbb{E}\int_0^\infty \frac{1}{r}\mathbf{1}\{M_{\boldsymbol{u}} \le r\} e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{u})} e^{-1/r} r^{-2}dr.$$

Noting that $\mathbb{E}\mathbf{1}\{M_{\boldsymbol{u}} \le r\} = \mathbb{P}\{\boldsymbol{X} \le r\boldsymbol{u}\} = e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{u})/r}$, applying Fubini's theorem to the last relation gives

$$\begin{aligned}\mathbb{E}G_{\boldsymbol{u}} = \int_0^\infty \frac{1}{r}\mathbb{E}\mathbf{1}\{M_{\boldsymbol{u}} \le r\} e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{u})/r} e^{-1/r} r^{-2}dr \\ = \int_0^\infty \frac{1}{r} e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{u})/r} e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{u})/r} e^{-1/r} r^{-2}dr = (2.41),\end{aligned}$$

as desired. Now, confirming (2.42) is achieved in similar fashion. We have

$$\mathbb{E}\{G_{\boldsymbol{u}}G_{\boldsymbol{w}}\} = \frac{\mathbb{E}\left\{\int_0^{(V_{\boldsymbol{\theta}_0}(\boldsymbol{u})+1)/M_{\boldsymbol{u}}} t_1 e^{-t_1}dt_1 \int_0^{(V_{\boldsymbol{\theta}_0}(\boldsymbol{w})+1)/M_{\boldsymbol{w}}} t_2 e^{-t_2}dt_2\right\}}{(V_{\boldsymbol{\theta}_0}(\boldsymbol{u}) + 1)^2 (V_{\boldsymbol{\theta}_0}(\boldsymbol{w}) + 1)^2}$$

substituting $t_1 = (V_{\boldsymbol{\theta}_0}(\boldsymbol{u}) + 1)/r$ and $t_2 = (V_{\boldsymbol{\theta}_0}(\boldsymbol{w}) + 1)/s$ into each integral above gives

$$\begin{aligned}\mathbb{E}\left\{\int_0^\infty \int_0^\infty \mathbf{1}\{M_{\boldsymbol{u}} \le r\}\mathbf{1}\{M_{\boldsymbol{w}} \le s\} e^{-(V_{\boldsymbol{\theta}_0}(\boldsymbol{u})+1)/r} e^{-(V_{\boldsymbol{\theta}_0}(\boldsymbol{w})+1)/s} r^{-3}s^{-3}drds\right\} \\ = \int_0^\infty \int_0^\infty \frac{1}{rs}\mathbb{E}\mathbf{1}\{\boldsymbol{X} \le r\boldsymbol{u}, \boldsymbol{X} \le s\boldsymbol{w}\} e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{u})/r} e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{w})/s} \omega_r(dr)\omega_r(ds)\end{aligned}$$

Note that $\mathbb{E}\mathbf{1}\{\boldsymbol{X} \leq r\boldsymbol{u}, \boldsymbol{X} \leq s\boldsymbol{w}\} = \mathbb{P}\left(\boldsymbol{X} \leq r\boldsymbol{u} \wedge s\boldsymbol{w}\right) = \exp\left(-V_{\boldsymbol{\theta}_0}\left(r\boldsymbol{u} \wedge s\boldsymbol{w}\right)\right)$. This implies

$$\int_0^\infty \int_0^\infty \frac{1}{rs} e^{-V_{\boldsymbol{\theta}_0}(r\boldsymbol{u} \wedge s\boldsymbol{w})} e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{u})/r} e^{-V_{\boldsymbol{\theta}_0}(\boldsymbol{w})/s} \omega_r(dr)\omega_r(ds)$$
$$= \int_0^\infty \int_0^\infty \frac{1}{rs} F_{\boldsymbol{\theta}_0}(r\boldsymbol{u} \wedge s\boldsymbol{w}) F_{\boldsymbol{\theta}_0}(r\boldsymbol{u}) F_{\boldsymbol{\theta}_0}(s\boldsymbol{w}) \omega_r(dr)\omega_r(ds) = (2.42),$$

and the proof is complete. □

CHAPTER 3

Hierarchical Gauss-Pareto models for spatial prediction of extreme precipitation

The max-stable models introduced in the previous Chapter have been deployed in a few recent works (Padoan et al., 2010; Davison et al., 2012; Thibaud et al., 2013) in order to characterize the dependence structure of extreme precipitation. However, numerous difficulties in dealing with max-stable models hamper their widespread use in practice, especially when spatial prediction is the goal (see e.g. Davison et al. 2012; Dombry et al., 2012; Wang and Stoev, 2011 and the references therein). Alternatively, generalized Pareto processes (Ferreira and de Haan, 2014) have emerged as a flexible class of spatial models for extremes. Such processes arise as limiting conditional distributions given a threshold exceedance (See Section 3.1 below for a precise definition), and thus are natural models for spatial prediction given that nearby observations are extreme.

In this Chapter, we propose a *Gauss-Pareto* process model for extreme precipitation which is closely related to the Pareto type models employed in recent manuscripts of Ferreira and de Haan (2014) and Thibaud and Opitz (2013). However, there exist key differences between our methodology and previous approaches based on threshold exceedances. First, we do not assume exact asymptotic distributions, rather our Gauss-Pareto model belongs to the max-domain of attraction of limiting max-stable

processes with spectral measures determined by an underlying Gaussian distribution (i.e. *spectrally Gaussian*). The advantage is that the model can be fit using standard MCMC methods for hierarchical models with latent Gaussian structure. This greatly simplifies inference while retaining the essential dependence characteristics of the most commonly used models for spatial extremes. A second key difference is the nature in which we handle partial censoring. While most precipitation measurements are essentially left-censored due to cumulative precipitation falling below reporting precision, when working with threshold exceedance models it is common to partially censor marginal observations that fall below a much higher threshold than those arising in data collection. Theoretical motivation for this lies in the fact that the model is derived asymptotically and thus including marginal observations that may not be approaching the asymptotic limit can lead to a poor approximation by the limit model (Smith, 1994; Coles, 2001 Section 8.3.1). On the other hand, we found it very common that 24 hour cumulative precipitation at various locations fall below such high thresholds even when nearby observations are extreme. This motivated us to consider a model that can account for such instances while maintaining essential tail dependence characteristics. We believe our methodology is new in this approach and allows us to consider larger spatial domains where there is greater chance of observing low cumulative precipitation given at least one extreme observation within the domain.

The rest of this Chapter is organized as follows: in the following Section 3.1 we review some basic motivating theory, highlighting connections between max-stable and Pareto processes. In Section 3.2 we define our model, give a detailed construction of the model hierarchy and specify the MCMC fitting procedure. The main application: spatial prediction of extreme summer precipitation in south central Sweden is presented in Section 3.3. Finally, we conclude with summary and directions for future work.

3.1 Pareto processes

Let $\eta = \{\eta(s)\}_{s\in S}$ be a non-negative stochastic process corresponding to a physical or environmental process over a compact spatial region of interest $S \subset \mathbb{R}^2$. Recall the characterization of max-stable processes as limiting distributions of scaled component-wise maxima

Condition 3.1. There exists a sequence of normalizing functions $a_n(s) > 0$ and $b_n(s)$ and a non-degenerate limit process $\zeta := \{\zeta(s)\}_{s\in S}$ for which the following holds

$$\left\{ \lim_{n\to\infty} \frac{\max_{i\leq n} \eta_i(s) - b_n(s)}{a_n(s)} \right\}_{s\in S} \overset{f.d.d.}{=} \{\zeta(s)\}_{s\in S}\,,$$

where η_i are independent copies of η and *f.d.d.* denotes equality in all finite dimensional distributions.

Hence the limit process ζ must be *max-stable.* For simplicity and without loss of generality (see Resnick, 1987 Prop 5.10(a)) we will assume that the max-stable process ζ is *simple max-stable,* i.e. ζ has identical margins that are standard Fréchet. In addition, we assume ζ has the following spectral representation

Proposition 3.2 (cf. Ferreira and de Haan, 2014, Prop. 2.3)**.** *Let $\zeta := \{\zeta(s)\}_{s\in S}$ be a simple max-stable process then*

$$\zeta(s) \overset{d}{=} \max_{i\in\mathbb{N}} \Gamma_i^{-1} V_i(s),$$

where $\{\Gamma_i\}_{i=1}^{\infty}$ are the points of a unit rate Poisson point process on $(0,\infty)$ and $V_i(s)$ are independent copies of a stochastic process V with $\mathbb{E}V(s) = 1$ and $\mathbb{E}\sup_{s\in S} V(s) < \infty$ a.s.

While max-stable models have been used recently in a variety of applications, as we noted in the prequel, inference suffers from a lack of tractable likelihoods (See e.g.

Einmahl et al., 2012). Furthermore, the complicated dependence structure imposed by taking point-wise maxima is often criticized as unrealistic because the point-wise maxima over a given period likely occurs at different times for different locations. This can obfuscate the true space-time dependence structure of the underlying phenomena η. Furthermore, conditional sampling (prediction) with spectrally Gaussian max-stable models is not straightforward and can be computationally prohibitive (Dombry et al., 2012; Wang and Stoev, 2011).

While max-stable processes and point-wise maxima can still be a useful framework, the underlying scientific motivation for our work lies in the case where one wants to characterize the dependence structure of η given that $\eta(s)$ is *large* for some s belonging to a finite collection $\{s_1, \ldots, s_d\}$ of observed locations. In statistical terms, this rather precise objective would be to characterize ρ_u a probability measure on $C^+(S)$ (the space of non-negative continuous functions on S) such that for a large threshold $u \gg 0$

$$P\left(\eta \in A | \max_{s \in \{s_1, \ldots, s_d\}} \eta(s) > u\right) = \rho_u(A), \tag{3.1}$$

for $A \in \mathcal{B}\left(C^+(S)\right)$ where $\mathcal{B}$ denotes the Borel sigma field. In practice, ρ_u can rarely be inferred directly, but it's characteristics can be approximated by limit distributions for the LHS of (3.1) in the following sense

Condition 3.3. There exists functions $a_u(s) > 0$ and $b_u(s)$, both continuous in s for every u, and for which the following holds

(i) For every $s \in S$, b_u is increasing in u.

(ii) $\lim_{u \to \infty} P\left(\eta(s) > b_u(s), \text{ for some } s \in S\right) = 0$.

(iii) There exists ρ, a non-degenerate probability measure on $C(S)$ such that for all

$A \in \mathcal{B}(C(S))$ with $\rho(\partial \mathrm{A})=0$

$$P\left(T_u(\eta) \in A | \eta(s) - b_u(s) > 0 \text{ for some } s \in S\right) \to \rho(A), \quad \text{as } u \to \infty \quad (3.2)$$

where

$$T_u(\eta) := \left\{ \frac{\eta(s) - b_u(s)}{a_u(s)} \right\}_{s \in S}.$$

The condition suggests that if $T_u(\eta) \in A | \eta(s) - b_u(s) > 0$ converges, then under judicious normalization we should approximate ρ_u by the limit measure ρ when u is large. The task then becomes characterizing the class of possible limits ρ, this is the main subject of Ferreira and de Haan (2014). For convenience we summarize the result

Proposition 3.4. *If Condition 3.3 holds, then there exists a sequence of normalizing transformations*

$$\widetilde{T}_u(\eta) := \left\{ \frac{\eta(s) - \tilde{b}_u(s)}{\tilde{a}_u(s)} \right\}_{s \in S}, \ \tilde{a}_u > 0,$$

such that for all $A \in \mathcal{B}(C(S))$

$$\lim_{u \to \infty} P\left(\tilde{T}_u(\eta) \in A | \sup_{s \in S} \left\{ \frac{\eta(s) - \tilde{b}_u(s)}{\tilde{a}_u(s)} \right\} > 1 \right) = P\left(ZX \in A\right), \quad (3.3)$$

where Z *is a Pareto random variable and* $X := \{X(s)\}_{s \in S}$ *is a non-negative stochastic process, independent of* Z*, with* $\mathbb{E}X(s) > 0$ *for all* $s \in S$ *and* $\sup_{s \in S} X(s) = c > 0$ *a.s.*

note. This is a direct result of Theorem 3.2, Condition 3.2, Corollary 3.1 and Theorem 2.1(3) of Ferreira and de Haan (2014). □

Remark 3.5 (cf. Example 3.2 of Ferreira and de Haan, 2014). Proposition 3.4 is equivalent to η belonging to the max-domain of attraction of a max-stable process

$\zeta := \{\zeta(s)\}_{s\in S}$ with spectral representation $\zeta \stackrel{d}{=} \max_{i\in\mathbb{N}} \Gamma_i^{-1} V_i$, where the $\{V_i\}_{i=1}^n$ are independent copies of ZX appearing in (3.3).

Proposition 3.4 suggests that processes of the form $V = ZX$, with $Z \sim GPD$ comprise a theoretically justified class of models for threshold exceedances which share dependence characteristics of popular max-stable models. Of the max-stable processes that have been proposed to model precipitation, most have spectral measure determined by an underlying Gaussian law (See e.g. Davison et al., 2012; Thibaud et al., 2013 and Chapter 2). These *spectrally Gaussian* models are max-stable attractors of the Pareto processes specified by $ZX(s) = Zf(W(s))$ where f is a continuous non-negative function and $W = \{W(s)\}_{s\in S}$, is a Gaussian process. This broad specification, however, is not always useful with respect to inference or prediction. The strategy we present here is to construct processes which share essential characteristics of popular max-stable and Pareto process models, yet remain amenable to straightforward MCMC techniques for Bayesian hierarchical models. For instance, we shall relax the requirement that $\sup_{s\in S} X(s) = c > 0$ a.s. To see why, consider the process $\bar{X} := X/\sup_{s\in S} X(s)$. Then $\sup_{s\in S} \bar{X}(s) = 1 > 0$ a.s., yet $\bar{X}$ retains the same spatial dependence structure as X. For convenience, we will call the processes $V(s) := Zf(W(s))$ *Gauss-Pareto* models. The reason for the Gaussian assumption is obvious in that it allows one to take advantage of the vast array of machinery developed for Gaussian processes including the specification of flexible parametric covariance functions, simulation of random fields and conditional sampling. Here we introduce two specifications of Gauss-Pareto models that belong to the max-domain of attraction of the spectrally Gaussian max-stable processes that have prevailed within the literature (See Section 2.1). These include the *Schlather* or *extremal Gaussian* type (Schlather, 2002) defined by

$$V^{(SC)}(s) := Z \max\{W(s), 0\},$$

and *Brown-Resnick* (Brown and Resnick, 1977; Kabluchko et al., 2009) type models

$$V^{(BR)}(s) := Z \exp\left(W(s) - \gamma(s)\right),$$

where γ is the semi-variogram of a centered intrinsically stationary Gaussian process W. The terms *Schlather model* and *Brown-Resnick model* refer to max-stable processes as defined in Remark 3.5. Nonetheless we adopt identical names here for their associated Gauss-Pareto processes. Two previous works using Schlather and Brown-Resnick max-stable models for precipitation extremes include Davison et al. (2012) where block maxima of summer precipitation near Zurich Switzerland is considered and Thibaud et al. (2013) who fit max-stable models to threshold exceedances of precipitation in the Val Ferret catchment, which also resides in Switzerland. Davison et al. (2012) found the Brown-Resnick type models dominated the Schlater models in terms of goodness of fit criterion. Conversely Thibaud et al. (2013) found that the Schlather model produced a better fit over the small Val Farret region. This is possibly due to the fact that there exists a lower bound on the range of extremal dependence at large lags under Schlather models as characterized by the *extremal coefficient* (Smith, 1990), a summary dependence measure that we study in further detail in Chapter 4. Indeed, our preliminary data analyses of extreme precipitation over south central Sweden (See Section 3.3 below) found that the Schlather type Gauss-Pareto model produced model fits that yielded overly strong spatial dependence between pairs of sites that were far apart. This led to spatial predictions that were biased towards heavier precipitation.

3.2 A log-Gauss-Pareto model for extreme precipitation

To develop our model, we began with an exploratory data analysis which revealed that the location and spatial range of extreme 24 hour cumulative precipitation over

our region of interest varied greatly from storm to storm. Thus, in order to make accurate spatial predictions, it is necessary to capture both the spatial range and profile of a particular extreme precipitation event as well as the approximate center. Consequently, we develop the following Brown-Resnick type model, which can be specified in a hierarchical manner to achieve the desired flexibility.

Definition 3.6. Let Z $\sim$ GPD(0,σ,ξ) and $W := \{W(s)\}_{s\in S}$ be a Gaussian process independent of Z. If

$$V(s) := Z \exp\{W(s)\}, \tag{3.4}$$

then $V := \{V(s)\}_{s\in S}$ a log-Gauss-Pareto process driven by W.

The following are immediate

1. $V(s)|W \sim \text{GPD}(0, \sigma e^{W(s)}, \xi)$, for all $s \in S$.
2. $Y|Z := \{\log(V(s))|Z\}_{s\in S}$ is a Gaussian process on S.

Without loss of generality, set $\sigma = 1$, otherwise replace $W(s)$ in Definition 3.6 with $W'(s) = W(s) + \log\sigma$. To capture varying location and range of 24 hour precipitation events, we consider processes driven by a *fractional Brownian surface* with drift. Specifically, we take the process W to be

$$W(s) = \varepsilon(s) + B(s) - (\|s - \omega\|/\lambda)^\alpha, \ \alpha \in (0, 2),$$

where $B := \{B(s)\}_{s\in S}$ is a mean-zero Gaussian process with $B(\omega) = 0$ almost surely for some origin point $\omega \in S$, and covariance function given by

$$K(s_1, s_2|\boldsymbol{\theta}) = \lambda^{-\alpha} \times \{\|s_1 - \omega\|^\alpha + \|s_2 - \omega\|^\alpha - \|s_1 - s_2\|^\alpha\} \tag{3.5}$$

with $\boldsymbol{\theta} := (\lambda, \alpha, \omega)$. Here $\|\cdot\|$ is Euclidean distance, λ determines range and α governs the smoothness of the process. Lastly, $\varepsilon(s)$ is a trend surface that captures spatially

varying scale. Hence our final model has the form

$$V(s) = Z \exp\left\{\varepsilon(s) + B(s) - (\|s - \omega\|/\lambda)^{\alpha}\right\}. \tag{3.6}$$

Note that the drift term $\gamma(s) := (\|s - \omega\|/\lambda)^{\alpha}$ is indeed the semi-variogram of the process B.

To interpret the model (3.6), consider $E(s) := \exp\{\varepsilon(s) + B(s) - (\|s - \omega\|/\lambda)^{\alpha}\}$ as the profile of extreme 24-hour cumulative precipitation which is roughly centered at ω. After controlling for scale within ε, the spatial maximum of the event profile is located, *near* the event center ω with high probability. Heuristically, this explained by the drift term $(\|s - \omega\|/\lambda)^{\alpha}$, which tends to infinity as one moves away from the origin ω. The intensity of the precipitation event is determined by Z, which exhibits the power law behavior that is characteristic of extreme precipitation. Simulated realizations from the log-Gauss-Pareto process driven by fractional Brownian surfaces are shown in Figure 3.1.

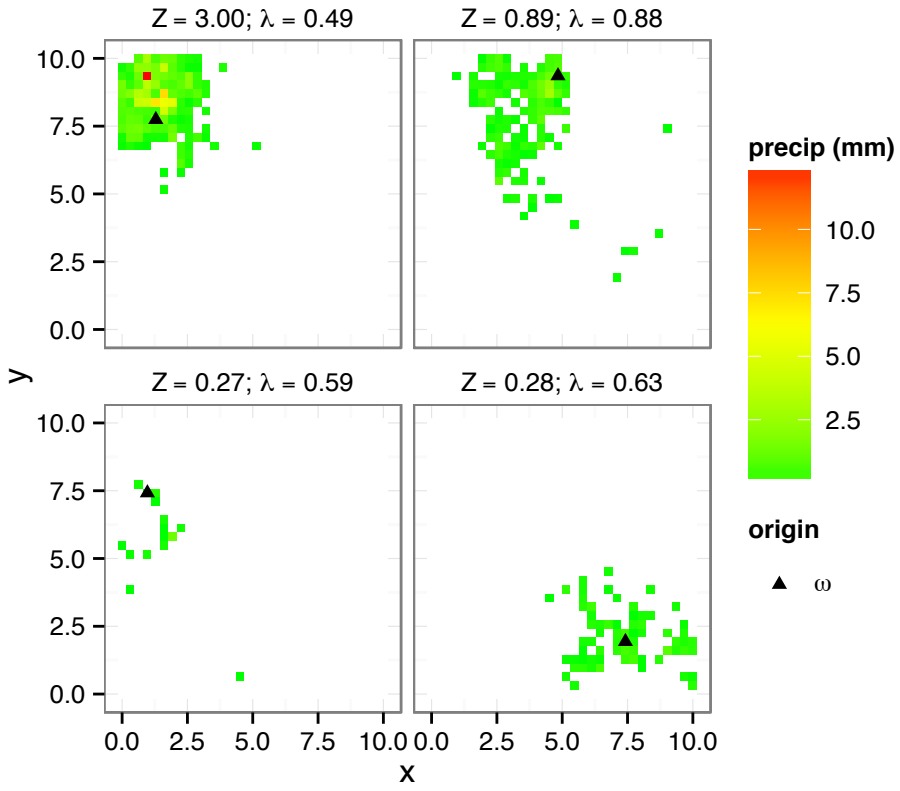
3.2.1 Model hierarchy

We now consider a series of independent extreme 24 hour precipitation events $V_i = \{V_i(s)\}_{s \in S}, i = 1, \ldots, n$. The V_i are identically distributed according to (3.6), each with strength Z_i, range λ_i and center ω_i. We will assume that the overall smoothness α, and trend surface ε, remain constant across 'time' $i = 1, \ldots, n$. At our disposal are measurements $V_i(s_j)$ at a sparse set of locations $s_1, \ldots, s_d \in S \subset \mathbb{R}^2$. It is convenient to write the log transformation of model (3.6)

$$Y_i(s) := \log V_i(s) = \log(Z_i) + \log\left(E_i(s)\right), \tag{3.7}$$

where $\log(E_i(s)) = \varepsilon(s) + B_i(s) - (\|s - \omega_i\|/\lambda_i)^{\alpha}$. Observe that the origin ω_i and scale λ_i, of the fractional Brownian surface B_i vary with each independent event.

Figure 3.1: Four realizations from the log-Gauss-Pareto process (3.6) with $Z \sim \text{GPD}(\mu = 0, \sigma = 1, \xi = 0.5)$, $\lambda \sim \exp(1)$, $\omega \sim \text{uniform}(S)$, and $\alpha = 0.5$. The process has been censored below 0.1mm. The ▲ correspond to the process origin ω for each of the four realizations.



Based on the the log transform (3.7), we formulate the model in three hierarchies, the *data generating level, process level,* and *prior.* Throughout, we use the notation $\boldsymbol{Y}_i = (Y_i(s_1), \ldots, Y_i(s_d))$ to denote the vector of log transformed observations and $\Sigma_{\boldsymbol{\theta}_i}$ to denote the $d \times d$ matrix with entries $\Sigma_{\boldsymbol{\theta}_i}(j,k) = K(s_j, s_k|\boldsymbol{\theta}_i)$, where K is the covariance function defined in (3.5). Similarly, we denote $\boldsymbol{\varepsilon} = (\varepsilon(s_1), \ldots, \varepsilon(s_d))$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta}_i} = (\gamma_i(s_1), \ldots, \gamma_i(s_d))$.

Data generation level

Let $\boldsymbol{\theta}_i' = (Z_i, \boldsymbol{\varepsilon})$ for each event $i \in \{1 \ldots, n\}$. Following the model (3.7), we have that

$$\boldsymbol{Y}_i|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i \sim N_d\left(\boldsymbol{\mu}_i, \Sigma_{\boldsymbol{\theta}_i}\right) \tag{3.8}$$

where $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{id})$ is a mean vector with elements

$$\mu_{ij} = \log(Z_i) + \varepsilon(s_j) - (\|s_j - \omega_i\|/\lambda_i)^\alpha.$$

In our application, censoring limits arise from the data collection process where observations below a threshold are not available due to reporting precision. For a given precipitation event $i \in \{1, \ldots, n\}$ we only observe the elements of $\boldsymbol{Y}_i$ that fall above a reporting threshold $l \in (-\infty, \infty)$. To be more concrete, we observe $\boldsymbol{Y}_{O_i}$ where $O_i \subset \{1, \ldots, d\}$ indicates the subset of observations from event i that occur above the threshold l whereas $\boldsymbol{Y}_{C_i}$ with $C_i = \{1, \ldots, d\} \backslash O_i$ are the censored data falling in the interval $(-\infty, l)$. Due to partial censoring, the observed information $\mathcal{D}_i$ for event i is

$$\mathcal{D}_i = \{Y_i(s_j); j \in O_i\} \cup \{Y_i(s_j) \leq l; j \notin C_i\}$$

Hence, the likelihood for the process $\boldsymbol{\theta}_i'$ given $\boldsymbol{\theta}_i, \mathcal{D}_i$ is

$$L(\boldsymbol{\theta}_i'; \boldsymbol{\theta}_i, \mathcal{D}_i) = p(\boldsymbol{Y}_{O_i}|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i) \int_{\boldsymbol{y} \leq l\mathbf{1}} p(\boldsymbol{y}|\boldsymbol{Y}_{O_i}, \boldsymbol{\theta}_i', \boldsymbol{\theta}_i) d\boldsymbol{y} \tag{3.9}$$

where $p(\boldsymbol{Y}_{O_i}|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i)$ and $p(\boldsymbol{y}|\boldsymbol{Y}_{O_i}, \boldsymbol{\theta}_i', \boldsymbol{\theta}_i)$ are multivariate Gaussian densities derived from (3.8). To handle integration in (3.9) we follow the augmentation method of De-Oliveira (2005) by embedding Monte-Carlo integration within our Bayesian MCMC. The exact algorithm is discussed in 3.2.2.

Process level

Here we specify the model for the process $\boldsymbol{\theta}_i' = (Z_i, \boldsymbol{\varepsilon})^\top$. By construction, we have that $\{Z_i\}_{i=1}^n \overset{iid}{\sim} \text{GPD}\,(0, 1, \xi)$. The large scale trend ε is modeled as a Gaussian process whose d-dimensional projection is multivariate normal: $\boldsymbol{\varepsilon} \sim N_d\,(\mathbf{0}, \sigma_\varepsilon^2 \Lambda_{\phi_\varepsilon})$, where $\Lambda_{\phi_\varepsilon}$ is a $d \times d$ correlation matrix with entries $\Lambda_{\phi_\varepsilon}(j,k) = \exp\,(-\|s_j - s_k\|/\phi_\varepsilon)$. Letting $\boldsymbol{\nu} = (\xi, \sigma_\varepsilon^2, \phi_\varepsilon)$, the density of the process model $\boldsymbol{\theta}_i'$ given $\boldsymbol{\nu}$ is

$$p(\boldsymbol{\theta}_i'|\boldsymbol{\nu}) \propto \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\boldsymbol{\varepsilon}^\top \Lambda_{\phi_\varepsilon}^{-1} \boldsymbol{\varepsilon}\right\} (1 + \xi Z_i)_+^{-1/\xi - 1}\,.$$

Prior

Prior distributions for the parameters $\boldsymbol{\nu} = (\xi, \sigma_\varepsilon^2, \phi_\varepsilon)$ and $\boldsymbol{\theta}_i = (\lambda_i, \alpha, \omega_i)$, are assumed to be independent. For σ_ϵ^2 we use the conjugate prior $\sigma_\varepsilon^2 \sim \text{IG}\,(a_\varepsilon/2, b_\varepsilon/2)$ where $\text{IG}\,(a, b)$ stands for the inverse Gamma distribution with mean $b/(a - 1)$. We also specify $\phi_\varepsilon \sim \text{IG}(c_\varepsilon, d_\varepsilon)$. For the shape parameter ξ we use a normal prior $\xi \sim N(m_\xi, v_\xi^2)$ and we let $\{\lambda_i\}_{i=1}^n \overset{iid}{\sim} \text{Gamma}\,(a_\lambda, \beta_\lambda)$ where a_λ and b_λ may be determined by initial hypotheses about the spatial extent of storms. Lacking a priori information about the center of extreme 24 hour precipitation events, we let $\{\omega_i\}_{i=1}^n \overset{iid}{\sim} \text{Uniform}(S)$. Lastly, α is assumed to be a fixed constant. In practice

it is difficult to estimate α unless the set of observations are very dense. For our analysis of Section 3.3, we fit the model by fixing $\alpha \in \{0.1, 0.2, \ldots, 1.9\}$ and select the corresponding α with the best fit as judged by goodness of fit criterion. Thus the prior distributions for $\boldsymbol{\nu}$ and $\boldsymbol{\theta}_i$ are given by

$$p(\boldsymbol{\nu}) \propto \exp\left\{-\frac{(\xi - m_\xi)^2}{2v_\xi^2}\right\} \left(\sigma_\varepsilon^2\right)^{-a_\varepsilon/2-1} \times \exp\left\{-\frac{b_\varepsilon}{2\sigma_\varepsilon^2}\right\} (\phi_\varepsilon)^{-c_\varepsilon-1} \exp\left\{-\frac{d_\varepsilon}{\phi_\varepsilon}\right\}$$

$$p(\boldsymbol{\theta}_i) \propto \lambda_i^{a_\lambda - 1} \exp\{-b_\lambda \lambda_i\}$$

3.2.2 MCMC sampling

For each $i \in \{1, \ldots, n\}$, the hierarchical model detailed above yields the following posterior density given the data $\mathcal{D}_i$

$$\begin{aligned} & p\left(\boldsymbol{Y}_{C_i}, \boldsymbol{\theta}_i', \boldsymbol{\theta}_i, \boldsymbol{\nu} | \mathcal{D}_i\right) \\ & \quad \propto p(\boldsymbol{\theta}_i' | \boldsymbol{\nu}) p(\boldsymbol{\theta}_i) p(\boldsymbol{\nu}) p(\mathcal{D}_i | \boldsymbol{Y}_{C_i}, \boldsymbol{\theta}_i', \boldsymbol{\theta}_i) p(\boldsymbol{Y}_{C_i} | \boldsymbol{\theta}_i', \boldsymbol{\theta}_i) \\ & \quad \propto p(\boldsymbol{\theta}_i' | \boldsymbol{\nu}) p(\boldsymbol{\theta}_i) p(\boldsymbol{\nu}) \\ & \qquad \times |\Sigma_{\boldsymbol{\theta}_i}|^{-1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)^\top \Sigma_{\boldsymbol{\theta}_i}^{-1} \left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i\right)\right\} \\ & \qquad \times \prod_{j \in C_i} \mathbf{1}_{\{Y_i(s_j) \leq l\}}, \end{aligned}$$

where $\boldsymbol{Y}_i = (\boldsymbol{Y}_{O_i}, \boldsymbol{Y}_{C_i})$ denotes the 'full' vector of log transformed measurements. The censored components $\{y_{ij}, j \in C_i\}$ of the vector $\boldsymbol{Y}_i$ are initialized at the censoring limit l and then sampled individually at each iteration of the MCMC from their respective full conditionals which are univariate truncated normal:

$$p\left(y_{ij}|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i, \boldsymbol{\nu}, \mathcal{D}_i, j \in C_i\right) = p\left(y_{ij}|\boldsymbol{Y}_{i(j)}, \boldsymbol{\theta}_i', \boldsymbol{\theta}_i, \boldsymbol{\nu}, , j \in C_i\right)$$
$$\propto N\left(\mu_{ij} + \boldsymbol{v}_{i(j)}^\top \Sigma_{\boldsymbol{\theta}_{i(j)}}^{-1}(\boldsymbol{Y}_{i(j)} - \boldsymbol{\mu}_{i(j)}), v_{ij} - \boldsymbol{v}_{i(j)}^\top \Sigma_{\boldsymbol{\theta}_{i(j)}}^{-1}\boldsymbol{v}_{i(j)}\right) \times \mathbf{1}_{\{y_{ij} \leq l\}}.$$

Here $\boldsymbol{Y}_{i(j)}$ is $\boldsymbol{Y}_i$ with the jth element removed, $\boldsymbol{v}_{i(j)}$ is the jth column of $\Sigma_{\boldsymbol{\theta}_i}$ with the jth element removed, $\Sigma_{\boldsymbol{\theta}_{i(j)}}$ is $\Sigma_{\boldsymbol{\theta}_i}$with the jth row and column removed and v_{ij} is the jth diagonal element of $\Sigma_{\boldsymbol{\theta}_i}$. This *data augmentation* for censored observations ensures that likelihood contributions given the data $\mathcal{D}_i$ follow (3.9). See DeOliveira (2005) for further details. Under the assumption that $\boldsymbol{Y}_i$ and $(\boldsymbol{Y}_{C_i}, \mathcal{D}_i)$ contain the same information, the full conditionals for $\boldsymbol{\varepsilon}$ and σ_ε^2 can be sampled directly

$$\begin{aligned} &p\left(\sigma_\varepsilon^2|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i, \xi, \phi_\varepsilon, \boldsymbol{Y}_{C_i}, \mathcal{D}_i, i = 1, \ldots, n\right) \\ &\quad = p\left(\sigma_\varepsilon^2|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i, \xi, \phi_\varepsilon, \boldsymbol{Y}_i, i = 1, \ldots, n\right) \\ &\quad = \text{IG}\left(\frac{1}{2}(d + a_\varepsilon), \frac{1}{2}(b_\varepsilon + \boldsymbol{\varepsilon}^\top \Lambda^{-1}\boldsymbol{\varepsilon})\right), \end{aligned}$$

$$\begin{aligned} &p\left(\boldsymbol{\varepsilon}|Z_i, \boldsymbol{\theta}_i, \boldsymbol{\nu}, \boldsymbol{Y}_{C_i}, \mathcal{D}_i, i = 1, \ldots, n\right) \\ &\quad = p\left(\boldsymbol{\varepsilon}|Z_i, \boldsymbol{\theta}_i, \boldsymbol{\nu}, \boldsymbol{Y}_i, i = 1, \ldots, n\right) \\ &\quad = N_d\left(\tilde{\boldsymbol{\mu}}_\varepsilon, \widetilde{\Lambda}_\varepsilon\right), \end{aligned}$$

where $\tilde{\boldsymbol{\mu}}_\varepsilon = \sum_{i=1}^n \widetilde{\Lambda}_\varepsilon \Sigma_{\boldsymbol{\theta}_i}^{-1}(\boldsymbol{Y}_i + \boldsymbol{\gamma}_{\boldsymbol{\theta}_i} - \log Z_i \mathbf{1})$, $\widetilde{\Lambda}_\varepsilon = \Lambda_{\phi_\varepsilon}^{-1} + \sum_{i=1}^n \Sigma_{\boldsymbol{\theta}_i}^{-1}$ and $\boldsymbol{\gamma}_{\boldsymbol{\theta}_i} = \text{diag}\left(\Sigma_{\boldsymbol{\theta}_i}\right)/2$. The remaining full conditionals are non-standard and require Metropolis Hastings sampling. Letting

$$p(\boldsymbol{Y}_i|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i) = N_d\left(\boldsymbol{\mu}_i, \Sigma_{\boldsymbol{\theta}_i}\right),$$

we have

$$
\begin{aligned}
&p\left(Z_i|\boldsymbol{\varepsilon}, \boldsymbol{\theta}_i, \boldsymbol{\nu}, \mathbf{Y}_{C_i}, \mathcal{D}_i\right) \\
&\qquad \propto p\left(Z_i|\boldsymbol{\varepsilon}, \boldsymbol{\theta}_i, \mathbf{Y}_i\right) \propto \left(1+\xi Z_i\right)_+^{-1/\xi-1} p(\mathbf{Y}_i|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i).
\end{aligned}
$$

$$
\begin{aligned}
&p\left(\xi|Z_i, \boldsymbol{\theta}_i, \boldsymbol{\nu}, \mathbf{Y}_{C_i}, \mathcal{D}_i, i=1,\ldots,n\right) \\
&\qquad \propto p\left(\xi|Z_i, i=1,\ldots,n\right) \\
&\qquad \propto \left\{\prod_{i=1}^{n}\left(1+\xi Z_i\right)_+^{-1/\xi-1}\right\} \exp\left\{-\frac{(\xi-m_\xi)^2}{2v_\xi^2}\right\}.
\end{aligned}
$$

$$
\begin{aligned}
&p(\phi_\varepsilon|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i, \xi, \sigma_\varepsilon^2, \mathbf{Y}_{C_i}, \mathcal{D}_i, i=1,\ldots,n) \\
&\qquad \propto p(\phi_\varepsilon|\boldsymbol{\varepsilon}, \sigma_\varepsilon^2) \\
&\qquad \propto |\Lambda_{\phi_\varepsilon}|^{-1/2} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2}\boldsymbol{\varepsilon}^\top \Lambda_{\phi_\varepsilon}^{-1}\boldsymbol{\varepsilon}\right\} \\
&\qquad\qquad \times \left(\phi_\varepsilon\right)^{-c_\varepsilon-1} \exp\left\{-\frac{d_\varepsilon}{\phi_\varepsilon}\right\}.
\end{aligned}
$$

$$
\begin{aligned}
&p(\lambda_i|\boldsymbol{\theta}_i', \alpha, \omega_i, \mathbf{Y}_{C_i}, \mathcal{D}_i) \\
&\qquad = p(\lambda_i|\boldsymbol{\theta}_i', \alpha, \omega_i, \mathbf{Y}_i) \\
&\qquad \propto \lambda_i^{a_\lambda-1} \exp\left\{-b_\lambda\lambda_i\right\} p(\mathbf{Y}_i|\boldsymbol{\theta}_i', \boldsymbol{\theta}_i).
\end{aligned}
$$

$$p(\omega_i|\boldsymbol{\theta}', \lambda_i, \alpha, \boldsymbol{Y}_{C_i}, \mathcal{D}_i)$$
$$= p(\omega_i|\boldsymbol{\theta}', \lambda_i, \alpha, \boldsymbol{Y}_i) \propto p(\boldsymbol{Y}_i|\boldsymbol{\theta}'_i, \boldsymbol{\theta}_i).$$

For most parameters, we use normal proposals. For the non-negative parameters λ_i and ϕ_ε, proposals are made with $\log \lambda'_i \sim N(\log \lambda_i, \psi_\lambda)$ and $\log \phi'_\varepsilon \sim N(\log \phi_\varepsilon, \psi_\varepsilon)$, where the candidate values $\lambda'_i, \phi'_\varepsilon$ are accepted with probability

$$\max\left\{0, \frac{p(\lambda'_i|\boldsymbol{\theta}'_i, \alpha, \omega_i, \boldsymbol{Y}_{C_i}, \mathcal{D}_i) \log \lambda'_i}{p(\lambda_i|\boldsymbol{\theta}'_i, \alpha, \omega_i, \boldsymbol{Y}_{C_i}, \mathcal{D}_i) \log \lambda_i}\right\},$$

and

$$\max\left\{0, \frac{p(\phi'_\varepsilon|\boldsymbol{\theta}'_i, \boldsymbol{\theta}_i, \xi, \sigma^2_\varepsilon, \boldsymbol{Y}_{C_i}, \mathcal{D}_i, i = 1, \ldots, n) \log \phi'_\varepsilon}{p(\phi_\varepsilon|\boldsymbol{\theta}'_i, \boldsymbol{\theta}_i, \xi, \sigma^2_\varepsilon, \boldsymbol{Y}_{C_i}, \mathcal{D}_i, i = 1, \ldots, n) \log \phi_\varepsilon}\right\}.$$

Normal proposal distributions are also used for ξ and Z_i, taking care that the initialized values for ξ and $\{Z_i\}_{i=1}^N$ satisfy $\min_{i=1\ldots,n}\{1 + \xi Z_i\} \geq 0$. Any proposals $\xi' \sim N(\xi, \psi_\xi), Z'_i \sim N(Z_i, \psi_Z)$ outside of the support for their respective full conditionals are accepted with nil probability. In the application below, proposals for the origin parameters ω_i were uniform on S which worked well, but a normal proposal could also be used.

3.2.3 Spatial prediction

Recall that our primary goal is to make predictions for $\{V_i(s)\}_{s\in S}$ given observations $\{\mathcal{D}_i\}_{i=1}^n$. In practice, posterior predictive distributions can be produced by sampling from the predictive distribution along a finite lattice $\widetilde{S} = \{\tilde{s}_1, \ldots, \tilde{s}_m\}$, conditional on $\boldsymbol{Y}_i, \boldsymbol{\theta}'_i, \boldsymbol{\theta}_i$. In order to do this, we must first sample the scale trend $\tilde{\boldsymbol{\varepsilon}} = (\varepsilon(\tilde{s}_1), \ldots, \varepsilon(\tilde{s}_m)) \in \mathbb{R}^m$ at prediction sites conditioned on the vales $\boldsymbol{\varepsilon}$ at the observation sites. The distribution of the Gaussian process model for the scale surface

ε at both the prediction and observation sites is

$$\begin{pmatrix} \tilde{\boldsymbol{\varepsilon}} \\ \boldsymbol{\varepsilon} \end{pmatrix} | \phi_\varepsilon, \sigma_\varepsilon^2 \sim N_{m+d} \left(\mathbf{0}, \sigma_\varepsilon^2 \begin{pmatrix} \Lambda_m & \Lambda_{dm}^\top \\ \Lambda_{dm} & \Lambda_\varepsilon \end{pmatrix} \right),$$

where Λ_m, Λ_{dm} are the corresponding matrices generated from the correlation function $\exp(-\|s - s'\|/\phi_\varepsilon)$. Hence, we sample from the conditional distribution

$$\tilde{\boldsymbol{\varepsilon}} | \boldsymbol{\varepsilon}, \phi_\varepsilon, \sigma_\varepsilon^2 \sim N_m \left(\tilde{\boldsymbol{\mu}}_\varepsilon, \sigma_\varepsilon^2 \widetilde{\Lambda}_\varepsilon \right).$$

where $\widetilde{\Lambda}_\varepsilon = \Lambda_m - \Lambda_{dm}^\top \Lambda_\varepsilon^{-1} \Lambda_{dm}$ and $\tilde{\boldsymbol{\mu}}_\varepsilon = \Lambda_{dm}^\top \Lambda_\varepsilon^{-1} \boldsymbol{\varepsilon}$. Next, consider the distribution of the data generating model at the prediction and observation sites

$$\begin{pmatrix} \widetilde{\boldsymbol{Y}}_i \\ \boldsymbol{Y}_i \end{pmatrix} | \boldsymbol{\theta}_i', \boldsymbol{\theta}_i, \tilde{\boldsymbol{\varepsilon}} \sim N_{m+d} \left(\begin{pmatrix} \boldsymbol{\mu}_{i,m} \\ \boldsymbol{\mu}_i \end{pmatrix}, \begin{pmatrix} \Sigma_{\boldsymbol{\theta}_i,m} & \Sigma_{\boldsymbol{\theta}_i,dm}^\top \\ \Sigma_{\boldsymbol{\theta}_i,dm} & \Sigma_{\boldsymbol{\theta}_i} \end{pmatrix} \right),$$

where $\boldsymbol{\mu}_{i,m}$ is a vector with entries $\mu_{ij,m} = \log Z_i + \varepsilon(s_j) - (\|\tilde{s}_j - \omega_i\|/\lambda_i)^\alpha$ and $\Sigma_{\boldsymbol{\theta}_i,m}, \Sigma_{\boldsymbol{\theta}_i,dm}$ are the matrices generated from the covariance function $K(s, s'|\boldsymbol{\theta}_i)$. Consequently, we sample from the conditional distribution

$$\tilde{\boldsymbol{y}}_i | \boldsymbol{Y}_i, \boldsymbol{\theta}_i', \boldsymbol{\theta}_i, \tilde{\boldsymbol{\varepsilon}} \sim N_m \left(\tilde{\boldsymbol{\mu}}_i, \widetilde{\Sigma}_{\boldsymbol{\theta}_i} \right),$$

where $\widetilde{\Sigma}_{\boldsymbol{\theta}_i} = \Sigma_{\boldsymbol{\theta}_i,m} + \Sigma_{\boldsymbol{\theta}_i,dm}^\top \Sigma_{\boldsymbol{\theta}_i}^{-1} \Sigma_{\boldsymbol{\theta}_i,dm}$ and

$$\tilde{\boldsymbol{\mu}}_i = +\Sigma_{\boldsymbol{\theta}_i,dm}^\top \Sigma_{\boldsymbol{\theta}_i}^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i).$$

The sampling is repeated for each iteration of the MCMC, yielding a posterior predictive distribution $\{\tilde{\boldsymbol{v}}_i^{(k)}\}_{k=1}^N := \{\exp(\tilde{\boldsymbol{y}}_i^{(k)})\}_{k=1}^N$. The resulting posterior predictive distributions can then be used to calculate point estimates from quantiles and uncertainty is based on the distribution spread.

3.3 Extreme Summer precipitation over southern Sweden

We apply our model to extreme Summer 24 hour precipitation totals gathered from synoptic stations of the Swedish Meteorological and Hydrological Institute. The data is open access and available at `http://www.smhi.se`. We focus our study to south central Sweden (below N 65° latitude) during the summer months June, July and August where precipitation falls almost exclusively as rain. Daily observations span years 1961-2011. We select only observations with highest quality control flags. With these data, non-zero precipitation below 0.1mm is reported as zero, hence we set the left-censoring limit to 0.1mm for all observations. Based on record length and completeness, 21 stations were selected as observation sites while an additional 21 sites with sparse observations were designated for validation. Figure 3.2 is a map of the region of interest and locations of the synoptic stations.

To designate observations that are extreme, with each day t in the record $\{1961 : 2011\}$ we first calculate $V_t^{\max} = \max_{s\in\{s_1,\ldots,s_d\}} V_t(s)$ where $\{s_1, \ldots, s_d\}$ is the set of $d = 21$ observation locations. Then we select dates t for which $V_t^{\max}$ exceeded the 95th percentile of daily maximum observations $\{V_t^{\max}\}_{t\in\{1961:2011\}}$. As expected, this subset of data displayed strong temporal dependence containing multiple clusters of consecutive dates. To rule out dependent observations we select the dates corresponding to the largest $V_t^{\max}$ within each cluster of consecutive days. This resulted in a final sample of $n = 59$ dates with no evidence of temporal dependence. There were no missing data in this sample.

We fit our model using the MCMC sampling scheme described in Section 3.2.2 with 60,000 iterations. Proposal variances $\psi_\lambda, \psi_\varepsilon, \psi_\xi, \psi_Z$ were tuned using preliminary runs such that acceptance rates were between .25 and .40 (Gelman et al., 1996). Convergence of preliminary runs were monitored using trace plots. Posterior distributions for $\sigma_\varepsilon^2, \phi_\varepsilon$ and λ were not sensitive to prior specification and for these parameters vague priors were used. The specification of hyper-parameters is given in Table 3.1.

Figure 3.2: Map of synoptic stations over south central Sweden.

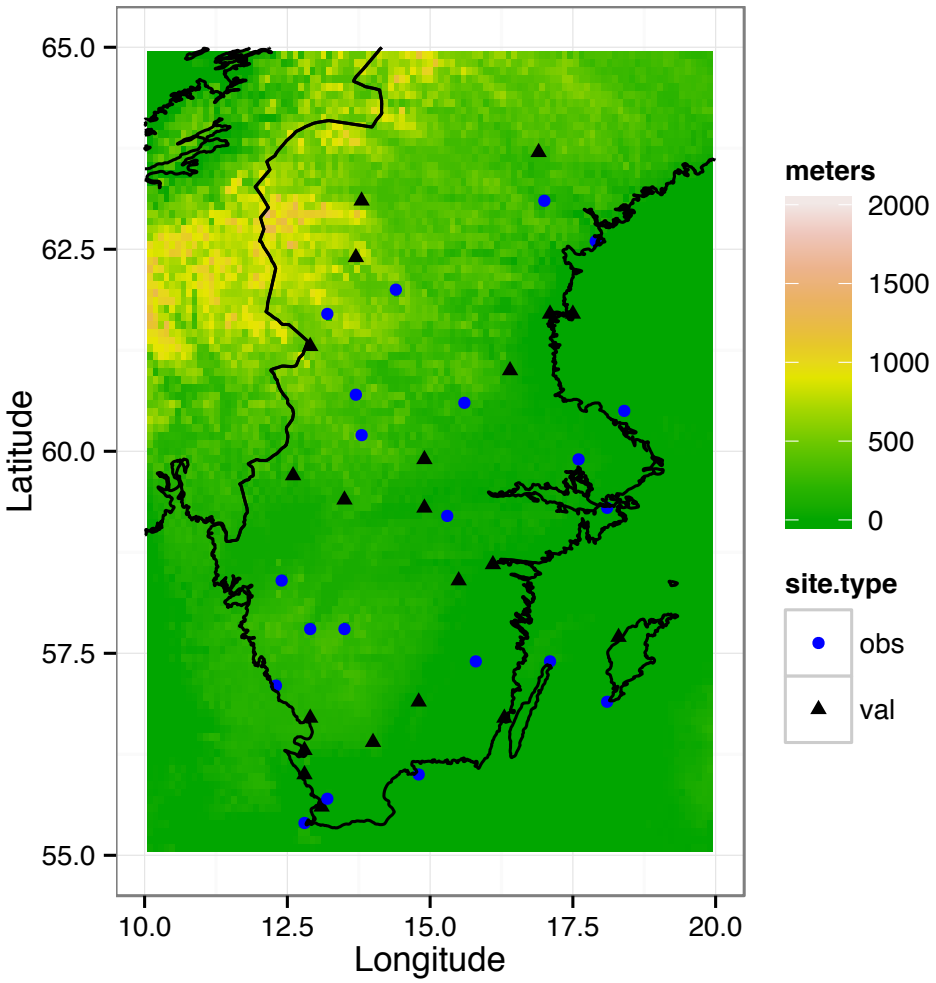


Table 3.1: Specification of hyper-parameters for MCMC.

σ_ε^2	ϕ_ε	ξ	λ
$a_\varepsilon = 5$ $b_\varepsilon = 2$	$c_\varepsilon = 3$ $d_\varepsilon = .25$	$m_\xi = .50$ $v_\xi^2 = .03$	$a_\lambda = .10$ $b_\lambda = .10$

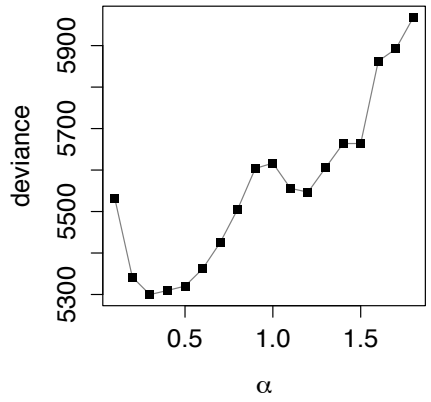
Preliminary MCMC runs indicated that the shape parameter ξ is difficult to estimate which is typical for spatial extremes (see e.g. Berrocal et al., 2014; Thibaud et al., 2013). Hence, we chose an informative prior based on previous literature for extreme precipitation where analyses typically suggested $\xi \in (0, 1)$ (Thibaud et al., 2013; Davison et al., 2012). In the final MCMC, the first 10,000 iterations were discarded as burn-in and then further thinned every 50 iterations to reduce serial correlations of the ω_i. This resulted in a posterior sample of size $N = 1000$. Due to the difficulty in estimating α, multiple chains were run for each $\alpha \in \{0.1, 0.2, \ldots, 1.9\}$. Because the number of parameters does not change as we vary α, we can compare model fit directly using the negative log-likelihood (deviance) at the data generating level

$$D = -2 \sum_{i=1}^{n} \log \left\{ \sum_{k=1}^{N} p(\boldsymbol{Y}_i^{(k)} | \hat{\boldsymbol{\theta}}_i', \hat{\boldsymbol{\theta}}_i) \right\}, \tag{3.10}$$

where $\boldsymbol{Y}_i^{(k)} = (\boldsymbol{Y}_{O_i}, \boldsymbol{Y}_{C_i}^{(k)})$, represent MCMC draws for censored observations concatenated with uncensored observations and $\hat{\boldsymbol{\theta}}_i', \hat{\boldsymbol{\theta}}_i$ denote posterior means of the MCMC sample for $\boldsymbol{\theta}_i'$ and $\boldsymbol{\theta}_i$. Note that (3.10) is derived from (3.9) and lower deviance scores correspond to better fit.

Figure 3.3 implies that the best model in terms of the deviance score corresponds to $\alpha = 0.3$. Hence, the remainder of the results are shown for $\alpha = 0.3$. Overall, it was difficult to detect significant differences in results for $\alpha \in (0.1, 1]$. Varying $\alpha \in (0.1, 1]$ had little effect on predictive performance and while the value of the range parameter λ_i adjusted accordingly, posterior distributions of the remaining pa-

Figure 3.3: Deviance scores (3.10) of model fit versus smoothness parameter α.



rameters did not noticeably change. We did notice that performance deteriorated for $\alpha > 1$. The MCMC procedure yields posterior point estimates of the event centers $\hat{\omega}_i := N^{-1} \sum_{k=1}^{N} \omega_i^{(k)}$. As an additional diagnostic, we examine the concordance between the estimated event center $\hat{\omega}_i$ and location of observed spatial maxima $s_i^{\max} := \arg\max_{s \in \{s_1, \ldots, s_d\}} V_i(s)$. While the true spatial maxima and even center ω_i are unknown, one would expect the observed $s_i^{\max}$ to be 'close' to $\hat{\omega}_i$. Indeed, Figure 3.4 supports a positive association between location of the observed maxima and the estimated event centers, particularly in the meridional (N-S) direction. The remaining fitted parameter estimates were constructed using posterior means. Specifically, $\hat{\xi} = 0.29$, $\hat{\sigma}_\varepsilon^2 = 3.70$ and $\hat{\phi}_\varepsilon = 26.67$. The distributions of the point estimates $\{\hat{\lambda}_i\}_{i=1}^n$ and $\{\hat{Z}_i\}_{i=1}^n$ are displayed in Figure 3.6.

The MCMC algorithm generates posterior predictive distributions

$$\hat{F}_{\tilde{s}}^{(i)}(v) := N^{-1} \sum_{k=1}^{N} \mathbf{1}_{\{v \leq v_i^{(k)}(\tilde{s})\}},$$

where $v_i^{(k)}(\tilde{s})$ represents the MCMC draw from the predictive distribution at location $\tilde{s}$ for iteration k. We evaluate the predictive distribution using the probability integral transform (PIT). If the distribution $\hat{F}_{\tilde{s}}^{(i)}$ is ideal, then probability integral transforms

Figure 3.4: Posterior mean of origin centers $\hat{\omega}_i$ versus location of maximum observation $s_i^{\max}$

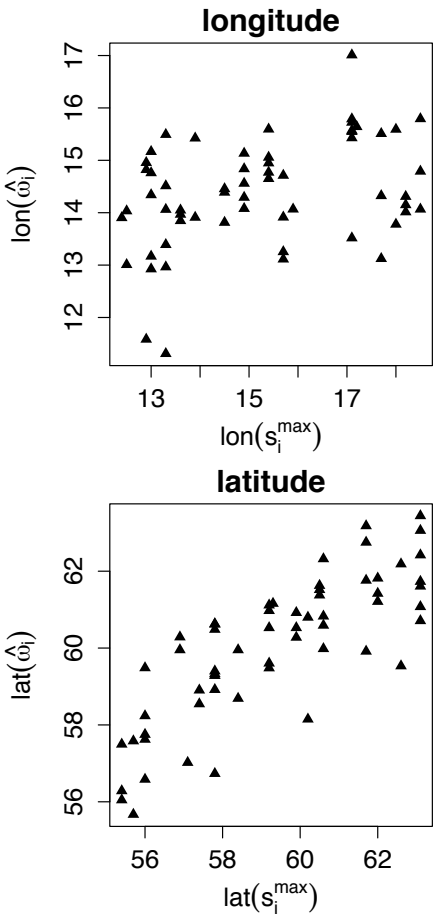


Figure 3.5: Distribution of posterior means for the range λ_i and intensities Z_i for each fitted date $i = 1, \ldots, 59$.

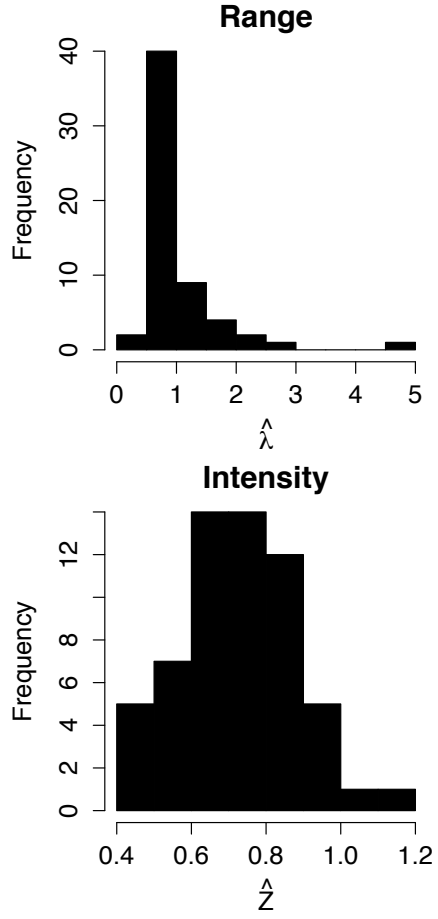
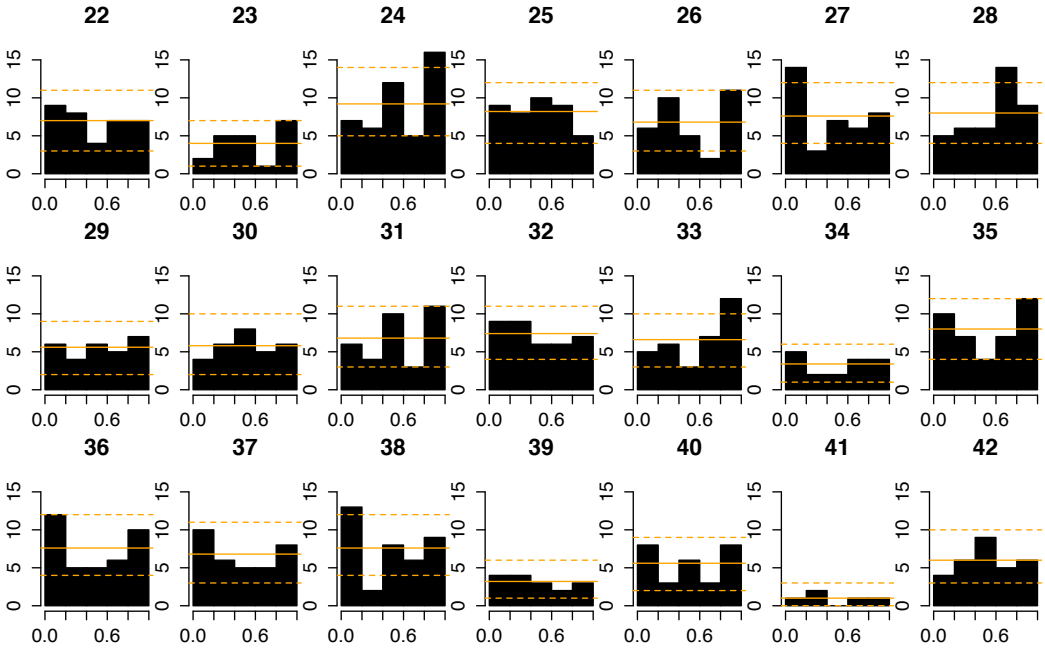


Figure 3.6: Probability integral transform histograms of the predictive distributions at 21 validation sites . Solid horizontal lines correspond to perfect uniformity. Confidence bands (dashed lines) are provided for reference, bin heights from random draws of a standard uniform distribution should fall outside of the dashed lines in approximately ten percent of cases.



$\{\hat{F}^{(i)}_{\tilde{s}}(V_i(\tilde{s})\}_{i=1}^n$ should be uniformly distributed on $(0, 1)$ (See e.g. Gneiting et al., 2007). Figure 3.6 displays the PIT histograms evaluated for the 21 validation sites. One should expect a reasonable amount of variance from the solid horizontal line indicating perfect uniformity. For reference, confidence bands indicating approximate 90 percent confidence intervals of the bar heights are shown. While some locations display a slight 'U' shape indicating underdispersion, overall the predictive distributions appear skillful when compared with random samples from a standard uniform distribution.

3.4 Discussion

We introduced Gauss-Pareto processes as a flexible class of models for extreme precipitation that can be fit using standard MCMC techniques for Bayesian hier-

archical modeling while retaining essential non-trivial dependence characteristics of popular max-stable and Pareto process models. Unlike the models used in Davison et al., 2012; Thibaud et al., 2013; Sang and Gelfand, 2009, spatial prediction is straightforward via conditional sampling from latent Gaussian processes and the hierarchical structure allows one to consider larger areas of interest by allowing the center and range of precipitation events to vary from storm to storm. Predictive distributions validated at holdout locations appear skillful and simulations from the process also appear realistic. There have been a very limited number of works for spatial prediction of extreme precipitation in the strict sense that we introduce in Section 3.1. and comparison with the alternative methods described in Ferreira and de Haan (2014) and Thibaud and Opitz (2013) are still pending, nonetheless our results appear promising.

Our methodology can be used for several applications in climate modeling such as fusion of climate model output with observational data for downscaling and testing distributional concordance between observed extremes and climate models, these are considerations for future work. Another interesting extension is motivated by the fact that extreme 24 hour precipitation displays strong temporal dependence. Consider the spatio-temporal model

$$V(s,t) = Z(t)\exp\left(W(s,t)\right)$$

where W is now a spatio-temporal Gaussian process and Z is a Heavy-tailed time dependent process. Fitting a spatio-temporal model would allow the borrowing of information from time-dependent observations leading to the incorporation of more data, although characterizing the dependence structure of Z is more challenging. A spatio-temporal model allows probabilistic forecasts of future precipitation given past observations that are extreme which is perhaps of greater interest than the purely

spatial prediction we consider in this Chapter.

CHAPTER 4

Tawn-Molchanov random vectors and bounding Value-at-Risk for the maximum loss

The asymptotic dependence of market risks plays a critical role in determining minimum capital requirements for financial institutions (Embrechts et al., 2009). In this Chapter, we propose a methodology for estimating upper bounds on probabilities of large losses, while accounting for the extremal dependence of market risks.

Suppose that the random vector $\boldsymbol{R} = (R_j)_{j=1}^d$, modeling portfolio losses, belongs to the max–domain of attraction of a max-stable random vector $\boldsymbol{V} \in \mathbb{R}^d$. That is, there exist $\boldsymbol{a}_n = (a_{n,i})_{i=1}^d \in \mathbb{R}_+^d$ and $\boldsymbol{b}_n = (b_{n,i})_{i=1}^d \in \mathbb{R}^d$, such that, as $n \to \infty$,

$$\left(\max_{k=1,\ldots,n} \frac{R_{k,i} - b_{n,i}}{a_{n,i}} \right)_{i=1}^d \xrightarrow{d} \boldsymbol{V} = (V_i)_{i=1}^d, \tag{4.1}$$

where $\boldsymbol{V}$ is a non–trivial random vector and $\boldsymbol{R}_k = (R_{k,i})_{i=1}^d,\ k = 1, \ldots, n$ are independent copies of $\boldsymbol{R}$. It is well known that the max–stable vector $\boldsymbol{V}$ retains essential extremal dependence characteristics of $\boldsymbol{R}$. Indeed, suppose that the marginal cumulative distribution functions $F_i(x) = \mathbb{P}(R_i \leq x),\ x \in \mathbb{R},\ i = 1, \cdots, d$ are *continuous*. Then, by (4.1) for the *coefficient of tail dependence* (Coles et al., 1999) between assets

i and j, we have

$$\begin{aligned}\lambda_{\boldsymbol{R}}(i,j) &:= \lim_{x\to\infty} \mathbb{P}\left(\frac{1}{1-F_i(R_i)} > x \,\Big|\, \frac{1}{1-F_j(R_j)} > x\right) \\ &= \lim_{x\to\infty} \mathbb{P}\left(\frac{1}{1-G_i(V_i)} > x \,\Big|\, \frac{1}{1-G_j(V_j)} > x\right) =: \lambda_{\boldsymbol{V}}(i,j), \qquad (4.2)\end{aligned}$$

where G_i denote the marginal cumulative distribution functions of $\boldsymbol{V}$. Note that $\lambda_{\boldsymbol{R}}(i,j) = \lambda_{\boldsymbol{R}}(j,i)$ can be interpreted as the probability that the losses R_i and R_j are extreme simultaneously, due to a common shock–factor affecting both (and perhaps larger group of) assets. More generally, from the perspective of multivariate regular variation, Relation (4.1) implies that $\boldsymbol{R}$ and $\boldsymbol{V}$ are tail–equivalent, in the following sense

$$\mathbb{P}\left(\frac{1}{1-\boldsymbol{F}(\boldsymbol{R})} \in xA\right) \sim \mathbb{P}\left(\frac{1}{1-\boldsymbol{G}(\boldsymbol{V})} \in xA\right) \sim \nu(xA), \quad \text{as } x\to\infty, \qquad (4.3)$$

with $\boldsymbol{F}(\boldsymbol{R}) = (F_i(R_i))_{i=1}^d$ and similarly $\boldsymbol{G}(\boldsymbol{V}) = (G_i(V_i))_{i=1}^d$ where above asymptotic equivalence is valid for all Borel $A \subset \mathbb{R}_+^d$ that are bounded away from the origin, such that $\nu(A) > 0$ and $\nu(\partial A) = 0$, that is, A is a continuity set for the *exponent measure* $\nu(xA) \equiv x^{-1}\nu(A)$ of $\boldsymbol{V}$ (see e.g., Propositions 5.8 and 5.10(b) in Resnick (1987)). Therefore, asymptotically precise results on the tail behavior of $\boldsymbol{R}$ can be obtained in terms of the exponent or, equivalently, the spectral measure of $\boldsymbol{V}$. For example, tail behavior of functions f of $\boldsymbol{R}$ can be characterized via (4.3) by considering sets of the form $A_t = \{\boldsymbol{u} \in \mathbb{R}^d : f(\boldsymbol{u}) > t\}$ (see e.g., Barbe et al. (2006) about the tail behavior of sums). It is well–known that for models based on multivariate Gaussian copula (unless trivial), the coefficients $\lambda_{\boldsymbol{R}}(i,j)$ vanish for $i \neq j$. This asymptotic independence property makes such models blind to contagion effects arising from simultaneous extremes.

In this Chapter, we advocate using a specific finite–dimensional max–stable model

for extreme risks, which although potentially high–dimensional, can be effectively estimated from data. Following the convention of Strokorb and Schlather (2013), this model will be referred to as the Tawn–Molchanov (TM) model, crediting earlier works of Schlather and Tawn (2002) and Molchanov (2008). We show that the TM model associated with the max–stable vector $\boldsymbol{V}$ in (4.1) yields an optimal upper bound on extreme *Value-at-Risk for the maximum portfolio loss* functional, referred to as *VaR–max*. This fact, provides a partial answer to a question raised in Embrechts et al. (2014) on the need for bounds on VaR for various convex functions of the portfolio. The result is a simple consequence of the recent work of Strokorb and Schlather (2013), which in turn extends seminal contributions due to Coles and Tawn (1996), Schlather and Tawn (2002) and Molchanov (2008). In practice, fitting TM models to portfolio data may be also of independent interest since their coefficients quantify contagion due to the simultaneous occurrence of extreme losses.

The rest of the Chapter is structured as follows. In the next Section 4.1 we introduce the TM model and describe its fundamental stochastic dominance property. We also relate its parameters to the tail dependence coefficients in (4.2), leading to appealing interpretations of the TM models in practice and visualization statistics such as the *tail dependence graph*. In Section 4.2, we introduce asymptotically justified and tight upper bounds on *VaR for the maximum loss* in a portfolio (VaR–max). We also examine the accuracy of these upper bounds for finite samples using simulations. In Section 4.3, we discuss the statistical estimation of the TM model using a simple regression–based procedure. The performance of the resulting estimators is evaluated with further simulations. In Section 4.4, we illustrate our methodology with an application to industry portfolios. We discuss the temporal scaling of VaR–max in Section 4.5. We comment on some computational challenges for high-dimensional portfolios in Section 4.6.

4.1 The Tawn-Molchanov model

For convenience, and without loss of generality, we shall work with *simple max–stable vectors* (SMS) $\boldsymbol{X} = (X_i)_{i=1}^d$ with standard unit Fréchet marginals. That is, $\mathbb{P}(X_i \leq x) = e^{-1/x},\ x > 0,\ i = 1, \ldots, d$. We will show in Section 4.2 below, how by applying monotone transformations, our results extend to the case of portfolios with different, light and/or heavy–tailed marginal distributions.

Definition 4.1. A SMS vector $\boldsymbol{Y} = (Y_i)_{i=1}^d$ is said to follow the Tawn-Molchanov (TM) model, if

$$\mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{x}) = \exp\left\{ - \sum_{J=\{i_1,\cdots,i_k\}\subset\{1,\cdots,d\}} \beta_J \frac{1}{x_{i_1} \wedge \cdots \wedge x_{i_k}} \right\}, \tag{4.4}$$

where the above summation is over *all* non–empty subsets of indices $J \subset \{1, \cdots, d\}$, $\boldsymbol{x} = (x_i)_{i=1}^d \in \mathbb{R}_+^d$, and where $\beta_J \geq 0,\ J \subset \{1, \cdots, d\}$ is a collection of non–negative coefficients.

Since the marginals of a TM vector $\boldsymbol{Y}$ are assumed to be standard unit Fréchet, we necessarily have that

$$\sum_{J\,:\,i\in J\subset\{1,\cdots,d\}} \beta_J = 1, \quad \text{for all } i = 1, \cdots, d. \tag{4.5}$$

Any collection of non–negative β_J's satisfying the above constraint yields a different TM model.

Alternatively, we have the following stochastic representation:

$$\boldsymbol{Y} \stackrel{d}{=} \bigvee_{k=1}^{p} \beta_{J_k} \mathbf{1}_{J_k} Z_k, \tag{4.6}$$

where $\mathbf{1}_{J_k} = \big(\mathbb{I}(\{i\} \cap J_k \neq \varnothing)\big)_{i=1}^d \in \{0,1\}^d$, the Z_k's are independent standard unit Fréchet random variables, and by convention, the sets J_k are ordered lexicographically,

as follows:

$$J_1 = \{1\}, \cdots, J_d = \{d\},\ J_{d+1} = \{1, 2\}, \cdots, J_p = \{1, \cdots, d\}, \quad \text{with } p := 2^d - 1.$$

In the context of market losses, one can interpret the Z_k's as independent heavy-tailed shock factors to a portfolio of assets $\{1, \ldots, d\}$. The shocks lead to maximum losses $Y_1, \ldots, Y_d$ according to the factor loadings $\beta_{J_1}, \ldots, \beta_{J_p}$. Each of the factors Z_k acts upon a subset $J_k \subset \{1, \ldots, d\}$. For instance, Z_1 represents a shock that affects only asset 1, while Z_p corresponds to shocks affecting the entire portfolio. An example of how to interpret coefficients $\beta_{J_1}, \ldots, \beta_{J_7}$ for a portfolio of 3 assets is given in Figure 4.1(a).

The following proposition, provides the motivation behind the TM model.

Proposition 4.2 (Corollary 33 in Strokorb and Schlather (2013))**.** *For every simple max-stable random vector $\boldsymbol{X} \in \mathbb{R}_+^d$, there exists a TM random vector $\boldsymbol{Y} = \mathrm{TM}(\boldsymbol{X})$ such that for all $x \in [0, \infty)$*

$$\mathbb{P}(X_j \leq x, j \in J) = \mathbb{P}(Y_j \leq x, j \in J),\ \textit{for all } \varnothing \neq J \subset \{1, \cdots, d\}. \tag{4.7}$$

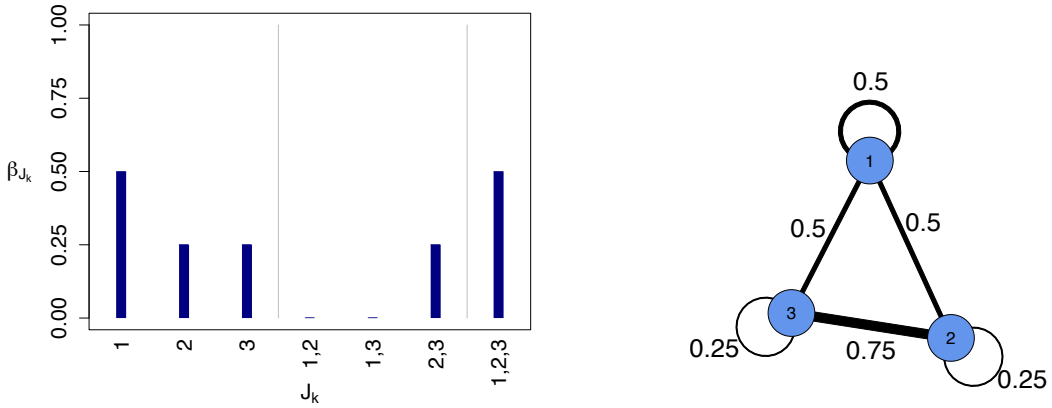
We have moreover that $\boldsymbol{X}$ is dominated by $\boldsymbol{Y} = \mathrm{TM}(\boldsymbol{X})$ in the lower orthant order, i.e.,

$$\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) \geq \mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{x}),\ \textit{for all } \boldsymbol{x} \in \mathbb{R}_+^d. \tag{4.8}$$

A proof of this result is given in Section 4.9. Relation (4.7) ensures that

$$\bigvee_{i \in J} X_i \overset{d}{=} \bigvee_{i \in J} Y_i, \quad \text{for all } J \subset \{1, \cdots, d\},$$

that is, the distributions of the maxima over all subsets of indices of $\boldsymbol{X}$ and $\mathrm{TM}(\boldsymbol{X})$ *match*. Intuitively, one can think of the distribution of $\boldsymbol{Y}$ as an estimate of the



(a) Coefficients of the TM model

(b) Tail dependence graph

Figure 4.1: (a) Interpreting $\boldsymbol{\beta}$ for a 3 asset portfolio: Since $\beta_{\{1\}} = 0.5$, half of the shocks to asset 1 occur independently from 2 and 3 and since $\beta_{\{1,2,3\}} = 0.5$, the remaining half of the shocks affecting 1 also affect assets 2 and 3. Similarly, since $\beta_{\{2\}} = 0.25$, one quarter of the shocks to asset 2 occur independently of 1 and 3; since $\beta_{\{2,3\}} = 0.25$, another quarter of the shocks to asset 2 occur independently of 1 but also affect 3, and since $\beta_{\{1,2,3\}} = 0.5$ the remaining half of the shocks affecting 2 also affect the entire portfolio. Finally, the sum of all β_{J_k} is equal to the extremal coefficient $\vartheta_{\boldsymbol{X}}(\{1,\ldots,d\}) = 1.75$. (b) Pairwise tail-dependence graph. Edge weights between nodes correspond to bivariate tail dependence coefficients $w(i,j)$ from (4.13). Edge weights at each node indicate level of asymptotic independence $\beta_{\{j\}}$ for components $j \in \{1,2,3\}$.

distribution of $\boldsymbol{X}$, under the specific constraint imposed by the structure of the TM model.

In Section 4.3, we will show that using (4.7), one can efficiently estimate the distribution of $\boldsymbol{Y}$ given independent realizations of $\boldsymbol{X}$, even when the distribution of $\boldsymbol{X}$ is unknown and itself difficult to estimate (arguably all pertinent cases). The property (4.8) is known as stochastic dominance in the lower orthant order , denoted $\boldsymbol{X} \leq_{\text{lo}} \boldsymbol{Y}$ (see Section 6.G.1 of Shaked and Shanthikumar, 2007). Consequently, by using the TM model associated with $\boldsymbol{X}$, we obtain stochastic upper bounds on all statistics of $\boldsymbol{X}$ whose ordering is preserved under (4.8). Some immediate questions that arise are: *What role does lower orthant dominance play in bounding classical measures of risk such as Value-at-Risk? How sharp are these bounds?* In Section 4.2, we show that the TM model yields an upper bound on VaR-max and illustrate the sharpness of this bound using simulations.

We conclude this section with several simple but illuminating properties of the coefficients of the TM model, which can help interpret them in terms of various measures of dependence. Let $\boldsymbol{X}$ be a simple max–stable random vector. Recall that the *extremal coefficient* $\vartheta_{\boldsymbol{X}}(J)$ of $\boldsymbol{X}$ corresponding to a non-empty subset of variables $J \subset \{1, \cdots, d\}$ is defined as follows

$$\mathbb{P}\left(\max_{j\in J} X_j \leq x\right) =: \exp\{-\vartheta_{\boldsymbol{X}}(J)/x\}, \quad x > 0. \tag{4.9}$$

That is, $\vartheta_{\boldsymbol{X}}(J)$ is the scale coefficient of the maximum of the components of $\boldsymbol{X}$ in J. It is well–known that

$$1 \leq \vartheta_{\boldsymbol{X}}(J) \leq |J|,$$

where the lower and upper bounds are achieved in the case when $X_j,\ j \in J$ are *completely* dependent or independent, respectively.

In view of (4.7), we have that

$$\vartheta_{\boldsymbol{X}}(J) = \vartheta_{\mathrm{TM}(\boldsymbol{X})}(J), \quad \text{for all } \varnothing \neq J \subset \{1, \cdots, d\},$$

that is, the *extremal coefficients* of $\boldsymbol{X}$ and its corresponding TM model *match*. Since the extremal coefficients do not generally determine the distribution of max–stable vectors, many simple max–stable models will correspond to the same TM model. Relation (4.4) readily implies that

$$\vartheta_{\boldsymbol{X}}(J) \equiv \vartheta_{\mathrm{TM}(\boldsymbol{X})}(J) = \sum_{K \cap J \neq \varnothing} \beta_K, \tag{4.10}$$

where the summation is over all sets $K \subset \{1, \cdots, d\}$ containing some element of J. One can also derive the inverse of (4.10) (see Theorem 4 of Schlather and Tawn, 2002)

$$\beta_K = \sum_{I \subset K} (-1)^{|I|+1} \vartheta_{\mathrm{TM}(\boldsymbol{X})}(K^c \cup I), \tag{4.11}$$

where $K^c = \{1, \ldots, d\} \backslash K$. This shows that the TM model is uniquely determined by either the set of extremal coefficients $\vartheta_{\mathrm{TM}(\boldsymbol{X})}(J)$ or the β_K coefficients.

Fix a set of variables $J = \{i_1, \cdots, i_k\} \subset \{1, \cdots, d\}$ and introduce the quantity

$$w_{\boldsymbol{X}}(J) := \sum_{K \,:\, J \subset K} \beta_K, \tag{4.12}$$

which is the sum of all coefficients in the TM model corresponding to sets K containing J. Note that by (4.5)

$$0 \leq w_{\boldsymbol{X}}(J) \leq (1 - \beta_{\{i_1\}}) \wedge \cdots \wedge (1 - \beta_{\{i_k\}}) \leq 1.$$

Proposition 4.13 in Section 4.8 shows that the $w_{\boldsymbol{X}}(J)$'s can be interpreted as the probability that the assets in J will be extreme simultaneously, that is, the cumulative

effect of all shocks Z_k that affect all assets in J. In the bivariate case $J = \{i, j\},\ i \neq j$ an interesting connection to the *coefficient of tail dependence* arises:

$$\lambda_{\boldsymbol{X}}(i,j) = \lambda_{\mathrm{TM}(\boldsymbol{X})}(i,j) = w_{\boldsymbol{X}}(i,j). \tag{4.13}$$

This implies that the coefficients of tail dependence of $\boldsymbol{X}$ and $\mathrm{TM}(\boldsymbol{X})$ coincide, which is another appealing feature of the TM model.

The bivariate tail dependence coefficients can be visualized through a weighted non–directed *tail dependence graph*, where the edge connecting nodes i and j is assigned weight $\lambda_{\boldsymbol{X}}(i,j) \equiv w_{\boldsymbol{X}}(i,j)$. The greater the weight, the stronger the extremal dependence between the variables. For completeness, one can also add an edge $i \leftrightarrow i$ with weight $\beta_{\{i\}}$, where the greater the weight, the greater the probability that shocks affecting only variable i occur. Figure 4.1(b) shows one example of such a dependency graph for the simple case of $d = 3$ assets.

4.2 The Tawn–Molchanov upper bound on VaR–max

In this section we establish concrete results that follow from the lower orthant dominance property of the TM model given in Proposition 4.2. We begin by defining VaR-f or the Value-at-Risk corresponding to quantiles of an arbitrary functional f of random losses:

Definition 4.3 (VaR-f)**.** Let $\boldsymbol{V} \in \mathbb{R}^d$ be a random vector representing extreme losses from a portfolio $\{1, \ldots, d\}$ after a fixed holding period. For a given function $f : \mathbb{R}^d \mapsto \mathbb{R}$ the *Value-at-Risk at level* $\alpha \in (0,1)$ is simply the α-quantile of the distribution of $f(\boldsymbol{V})$

$$\mathrm{VaR}_\alpha(f(\boldsymbol{V})) := \inf\left\{x : \mathbb{P}(f(\boldsymbol{V}) \leq x) > \alpha\right\}.$$

In this section we will focus on $f(\boldsymbol{V}) \equiv \boldsymbol{V}^{\vee} := \max_{j=1,\ldots,d} V_j$ which is the maximum loss of the portfolio and refer to $\mathrm{VaR}_{\alpha}(\boldsymbol{V}^{\vee})$ as *VaR–max*. In Chapter 5 we consider the more common *VaR–plus*: $f(\boldsymbol{V}) \equiv \boldsymbol{V}^{+} := \sum_{j=1}^{d} V_j$ representing the sum of losses.

Recall that an arbitrary max-stable random vector $\boldsymbol{V} = (V_j)_{j=1}^{d}$ has marginal distributions that are generalized extreme value (GEV) with location μ_j, scale $\sigma_j > 0$ and shape ξ_j, respectively for each component $j = 1, \ldots, d$. Standardization to 1-Fréchet marginals is achieved by applying the following transform

$$\boldsymbol{X} = \boldsymbol{T}(\boldsymbol{V}; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}) := \begin{pmatrix} T_1(V_1) = \left[1 + \xi_1 \left\{V_1 - \mu_1\right\} / \sigma_1\right]_{+}^{1/\xi_1} \\ \vdots \\ T_d(V_d) = \left[1 + \xi_d \left\{V_d - \mu_d\right\} / \sigma_d\right]_{+}^{1/\xi_d} \end{pmatrix}. \tag{4.14}$$

The transformation is monotone and thus admits an inverse

$$\boldsymbol{V} = \boldsymbol{T}^{-1}(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}) := \begin{pmatrix} T_j^{-1}(X_1) = \mu_1 + \frac{\sigma_1}{\xi_1}\left[(X_1)_{+}^{\xi_1} - 1\right] \\ \vdots \\ T_j^{-1}(X_d) = \mu_d + \frac{\sigma_d}{\xi_d}\left[(X_d)_{+}^{\xi_d} - 1\right] \end{pmatrix}. \tag{4.15}$$

It is well known (see Proposition 5.10 Resnick, 1987) that $\boldsymbol{T}$ and $\boldsymbol{T}^{-1}$ preserve the asymptotic dependence structure of $\boldsymbol{X}$ in $\boldsymbol{V}$ and visa-versa. When the random vector $\boldsymbol{V}$ has identical margins, we say that $\boldsymbol{V}$ is *homogeneous*.

Now were are ready to state the main result of this section.

Proposition 4.4. *Let $\boldsymbol{V} \in \mathbb{R}^d$ be an arbitrary max-stable random vector with GEV marginals determined by parameters $\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}$.*

(i) If

$$\boldsymbol{W} = \mathrm{TM}(\boldsymbol{V}) := \boldsymbol{T}^{-1}\left\{\mathrm{TM}\left[\boldsymbol{T}(\boldsymbol{V}; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi})\right]; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}\right\},$$

where $\boldsymbol{T}$ and $\boldsymbol{T}^{-1}$ are given in (4.14) *and* (4.15)*, then*

$$\mathrm{VaR}_\alpha(\boldsymbol{V}^\vee) \le \mathrm{VaR}_\alpha(\boldsymbol{W}^\vee). \tag{4.16}$$

(ii) If $\boldsymbol{V} = \mathrm{TM}(\boldsymbol{V})$*, then* (4.16) *holds with equality.*

(iii) If $\boldsymbol{V}$ *is homogeneous, i.e. the margins of* $\boldsymbol{V}$ *are identically distributed* $\mathrm{GEV}(\mu, \sigma, \xi)$*, then* (4.16) *holds with equality and*

$$\mathrm{VaR}_\alpha(\boldsymbol{V}^\vee) = \mu + \frac{\sigma}{\xi}\left[\left(\frac{\vartheta_{\mathrm{TM}(\boldsymbol{V})}(\{1,\dots,d\})}{-\log\alpha}\right)^{\xi} - 1\right]. \tag{4.17}$$

Proof. To prove (i) it suffices to show $\mathbb{P}(\boldsymbol{V}^\vee \ge x) \le \mathbb{P}(\boldsymbol{W}^\vee \le x)$, for all $x \in \mathbb{R}$. Let $(T_j)_{j=1}^d$ be the univariate monotone transformations defined in (4.14) and $(T_j^{-1})_{j=1}^d$ their inverses. Then it follows that

$$\begin{aligned}\mathbb{P}(\boldsymbol{V}^\vee \le x) = \mathbb{P}\left(\max_{j=1,\dots,d} V_j \le x\right) &= \mathbb{P}\left(T_j(V_j) \le T_j(x), j = 1,\dots,d\right) \\ &\overset{(4.8)}{\ge} \mathbb{P}\left(\mathrm{TM}(T(\boldsymbol{V}))_j \le T_j(x), j = 1,\dots,d\right) \\ &= \mathbb{P}\left(\max_{j=1,\dots,d} T_j^{-1}(\mathrm{TM}(T(\boldsymbol{V}))_j) \le x\right) = \mathbb{P}(\boldsymbol{W}^\vee \le x), \end{aligned} \tag{4.18}$$

where the inequality in (4.18) holds by Proposition 4.2. Part (ii) is trivial. To prove (iii), observe that under homogeneity

$$\begin{aligned}\mathbb{P}(\boldsymbol{V}^\vee \le x) &= \mathbb{P}(T_1(\boldsymbol{V}^\vee) \le T_1(x)) \\ &= \mathbb{P}(\boldsymbol{T}(\boldsymbol{V})^\vee \le T_1(x)) = \exp\left[-\vartheta_{\mathrm{TM}(\boldsymbol{V})}(\{1,\dots,d\})/T_1(x)\right], \end{aligned} \tag{4.19}$$

where the last equality follows from (4.9). The formula in (4.17) is obtained by inverting the last expression. □

We now comment on how to apply the above results in practice.

Remark 4.5. In reality, the vector of negative returns $\boldsymbol{R} = (R_i)_{i=1}^d$ is not precisely max–stable, but it is reasonable to assume that (4.1) holds. Let $\boldsymbol{R}_k,\ k = 1, \cdots, M$ be measurements of the negative returns $\boldsymbol{R}$ over M periods. Then, for sufficiently large M, the component–wise maximum

$$\boldsymbol{V}^{(M)} = \bigvee_{k=1}^{M} \boldsymbol{R}_k$$

is close in distribution to a max–stable vector. (This approximation is valid not only for independent $\boldsymbol{R}_k$'s but, in fact, under very general conditions on the temporal dependence of the returns, so long as they are stationary, cf Remark 1.1) We can thus apply Proposition 4.4 to obtain an estimate of an upper bound on VaR-max $\mathrm{VaR}_\alpha(\boldsymbol{V}^{(M),\vee})$. Naturally, by (4.1), this estimate becomes more accurate as M grows (see e.g., Proposition 0.1 in Resnick, 1987). On the other hand, as discussed in Section 4.5 below, for fixed M, the estimate also becomes more precise as α grows.

Remark 4.6. The quantity $\mathrm{VaR}_\alpha(\boldsymbol{V}^{(M),\vee})$ can be interpreted as the Value-at-Risk (at level α) for the maximum portfolio loss over the time period $[0, M]$. For example, if the losses are measured daily, the estimated $\mathrm{VaR}_\alpha(\boldsymbol{V}^{(M),\vee})$ corresponds to monthly, quarterly or yearly VaR–max if $M = 20,\ 60$, or 250.

On the other hand, one may be interested in quantifying *extreme VaR–max*, i.e., for values $\alpha \approx 1$, but over the original time–scale. In this case, $\mathrm{VaR}_\alpha(\boldsymbol{V}^{(M),\vee})$ – the M–period VaR–max – naturally overestimates $\mathrm{VaR}_\alpha(\boldsymbol{R}^\vee)$. We discuss simple methodology for *downscaling* $\mathrm{VaR}_\alpha(\boldsymbol{V}^{(M),\vee})$ in Section 4.5, below, under certain simplifying assumptions on the temporal dependence of the $\boldsymbol{R}_k$'s.

4.2.1 Example: VaR-max for simulated max-linear model

Here we provide a numerical study on the TM bounds for VaR-max established in Proposition 4.4. Specifically, we examine the behavior of the ratio

$$\frac{\mathrm{VaR}_\alpha(\boldsymbol{V}^\vee)}{\mathrm{VaR}_\alpha(\mathrm{TM}(\boldsymbol{V})^\vee)},\ \alpha \in (0,1) \tag{4.20}$$

when $\boldsymbol{V}$ is a max-stable random vector which does not follow a TM model. Proposition 4.4 asserts that the ratio (4.20) is bounded above by the value 1 for all $\alpha \in (0,1)$. In practice, it is important to know how sharp is the TM bound, in-particular for high quantiles ($\alpha \approx 1$). The following numerical study provides a qualitative answer to this question based on simulated data from the *max-linear* model.

Definition 4.7 (max-linear random vector)**.** Let $\boldsymbol{A} = (a_{jk})_{d\times p}$ be a matrix of non-negative entries such that $\boldsymbol{A}\mathbf{1} = \mathbf{1}$, i.e., with rows that sum to unity. Let $Z_1, \ldots, Z_p$ be a sequence of iid standard Fréchet random variables. A random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top \in \mathbb{R}^d_+$ with elements given by

$$X_j = \bigvee_{k=1}^{p} a_{jk} Z_k, \tag{4.21}$$

has standard Fréchet margins and is called *simple max-linear* (SML). It is easily verified that $\boldsymbol{X}$ has cumulative distribution function

$$\mathbb{P}\left(\boldsymbol{X} \leq \boldsymbol{x}\right) = \exp\left[-\sum_{k=1}^{p}\bigvee_{j=1}^{d}\frac{a_{jk}}{x_j}\right],\quad \boldsymbol{x} \in \mathbb{R}^d_+ \tag{4.22}$$

and extremal coefficient function $\vartheta_{\boldsymbol{X}}(J) = \sum_{k=1}^{p}\bigvee_{j\in J} a_{jk}$. Given arbitrary GEV pa-

rameters $\boldsymbol{\mu}, \boldsymbol{\sigma}$ and $\boldsymbol{\xi}$, one can construct (non-simple) max-linear random vectors via

$$\boldsymbol{V} = \boldsymbol{T}^{-1}(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}) = \begin{pmatrix} T_j^{-1}(X_1) = \mu_1 + \frac{\sigma_1}{\xi_1}\left[(X_1)_+^{\xi_1} - 1\right] \\ \vdots \\ T_j^{-1}(X_d) = \mu_d + \frac{\sigma_d}{\xi_d}\left[(X_d)_+^{\xi_d} - 1\right] \end{pmatrix}. \tag{4.23}$$

For this simulation, we considered two specifications of the coefficient matrix $\boldsymbol{A}$.

$$\boldsymbol{A}_1 = \begin{pmatrix} 0 & 1 \\ 1/4 & 3/4 \\ 1/2 & 1/2 \\ 3/4 & 1/4 \\ 1 & 0 \end{pmatrix}, \ \boldsymbol{A}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/2 & 1/10 & 1/10 & 1/10 & 1/10 & 1/10 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \end{pmatrix}. \tag{4.24}$$

These matrices were designed in such a way that the non-zero entries a_{jk} vary significantly within each column and therefore represent a substantial departure from the TM model. Indeed, the TM model requires that the non-zero coefficients a_{jk} in (4.22) for each fixed k, be constant in j (recall (4.6)). In addition, the number of underlying factors (2 and 6) are respectively less and greater than the dimension $d = 5$. This allows us to examine the effect of the number of independent shock factors on the complexity of the corresponding TM coefficients β_{J_k} (see Figures 4.3 and 4.4). To rule out the trivial case of homogeneity (recall Proposition 4.4(iii)), we considered three scenarios for the GEV parameters, *infinite variance*, *infinite mean* and *mixed tails*, shown in Table 4.1.

Because we are in the inhomogeneous setting, analytic expressions of the ratio (4.20) are not available and thus we employ a Monte-Carlo approximation. Specifically, we generate a large sample $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n$ from the max-linear model (4.23). Independently, we sample $\boldsymbol{W}_1 \ldots, \boldsymbol{W}_n \overset{iid}{\sim} \mathrm{TM}(\boldsymbol{V})$. By letting $\{\boldsymbol{V}_{(i)}^{\vee}\}_{i=1}^n$ and $\{\boldsymbol{W}_{(i)}^{\vee}\}_{i=1}^n$ be the order statistics of the individual component maxima and $\lfloor \cdot \rfloor$ the integer floor

Table 4.1: GEV parameter settings for VaR-max simulation studies

	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$	$\boldsymbol{\xi}$
mixed tails	$\mathbf{1}$	$\boldsymbol{\xi}$	$\left(\frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}, \frac{3}{2}\right)^\top$
infinite variance	$\mathbf{1}$	$\left(\frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}\right)^\top$	$\left(\frac{1}{2}\right) \times \mathbf{1}$
infinite mean	$\mathbf{1}$	$\left(\frac{1}{6}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{5}{6}\right)^\top$	$\mathbf{1}$

function, the ratio

$$\frac{\boldsymbol{V}^{\vee}_{(\lfloor n\alpha \rfloor)}}{\boldsymbol{W}^{\vee}_{(\lfloor n\alpha \rfloor)}} \approx \frac{\mathrm{VaR}_\alpha(\boldsymbol{V}^{\vee})}{\mathrm{VaR}_\alpha(\mathrm{TM}(\boldsymbol{V})^{\vee})} \tag{4.25}$$

is an empirical estimate of the ratio (4.20). We used a rather large sample size n =10,000, yet the variance of (4.25) near $\alpha \approx 1$ was considerably high. To compensate for the variability and in order to better understand the behavior of the ratio at extreme quantiles, we simulated 100 replications using sample size n =10,000 and examine the median of the resulting empirical ratios (4.25). Results are reported in Figure 4.2 where grey lines indicate 100 replicates of the ratio (4.25) and the dark line indicates the median of the 100 replicates. Observe that the median is essentially bounded above by the value 1, as expected by Proposition 4.4.

Note that even though the max-linear models we consider differ substantially from their TM counterparts, the TM upper bound provides relatively accurate estimates of VaR-max. The upper bound is tighter for the infinite variance scenario ($\boldsymbol{\xi} = (1/2) \cdot \mathbf{1}$) versus the infinite mean scenario ($\boldsymbol{\xi} = \mathbf{1}$). Interestingly, the bound appears asymptotically sharp (as $\alpha \uparrow 1$) for the mixed tails scenario.

4.3 Estimation for the Tawn-Molchanov model

By Relation (4.4), the distribution function of a TM random vector $\boldsymbol{Y}$ is fully specified by a *linear* function of the coefficients $\boldsymbol{\beta} = (\beta_{J_k})_{k=1}^p$, $p = 2^d - 1$, or equivalently extremal coefficients $\boldsymbol{\vartheta} = (\vartheta_{\boldsymbol{Y}}(J_k))_{k=1}^p$. Motivated by this fact, we develop

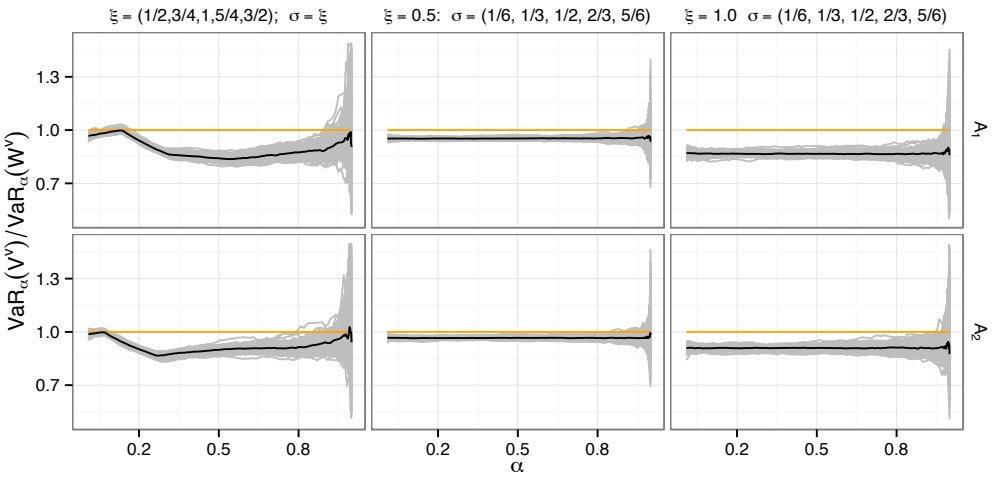


Figure 4.2: Ratio of Value-at-Risk for component-wise maxima of max-linear model $\boldsymbol{V}$ versus TM model $\boldsymbol{W} = \mathrm{TM}(\boldsymbol{V})$. Grey lines indicate 100 replicates of the ratio (4.25) and the dark line reports the median. Rows correspond to coefficient matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ in (4.24) for max-linear model (4.23). Columns indicate different values of the shape and scale parameters $\boldsymbol{\xi}$ and $\boldsymbol{\sigma}$ for the margins of $\boldsymbol{V}$.

a regression based estimation procedure. This circumvents much of the difficulties stemming from intractable likelihoods when working with general max-stable models. To begin, we introduce the following Lemma that establishes the connection between data generated from a max-stable random vector $\boldsymbol{X}$ and the parameters of its TM counterpart.

Lemma 4.8. *Let $\boldsymbol{X} \in \mathbb{R}_+^d$ be simple max-stable and $\boldsymbol{Y} = \mathrm{TM}(\boldsymbol{X})$ with coefficients $\beta_{J_1}, \ldots, \beta_{J_p}$.*

Then

$$\left(\max_{j \in J_k} X_j\right)^{-1} = \min_{j \in J_k} X_j^{-1},$$

is exponentially distributed with rate equal to the extremal coefficient

$$\vartheta_k := \sum_{\ell=1}^{p} \mathbb{I}(J_k \cap J_\ell \neq \varnothing)\beta_{J_\ell}, \quad 1 \leq \vartheta_k \leq |J_k|. \tag{4.26}$$

The proof follows from (4.9) and (4.10). Given $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \overset{iid}{\sim} \boldsymbol{X}$, Lemma 4.8 suggests an estimator for $\vartheta_1, \ldots, \vartheta_p$

$$\tilde{\vartheta}_k := \min\left\{1, \max\left\{|J_k|^{-1}, \frac{1}{n}\sum_{i=1}^{n} \min_{j \in J_k} X_{ij}^{-1}, \right\}\right\}^{-1}. \tag{4.27}$$

In the case $|J_k| = 2$, the estimator (4.27) has appeared in Smith (1990) and it is closely related to the estimator of the Pickands function (see e.g., Eq. (9.22) in Beirlant et al., 2004).

In general there is no guarantee that the estimated values $\tilde{\boldsymbol{\vartheta}} = (\tilde{\vartheta}_k)_{k=1}^p$ lie within the feasible region corresponding to a set of extremal coefficients of a valid probability distribution (see Corollary 5 in Schlather and Tawn, 2002). To correct for this, we propose to find $\boldsymbol{\vartheta}$ in the feasible region that minimizes the Euclidean distance between $\boldsymbol{\vartheta}$ and $\tilde{\boldsymbol{\vartheta}}$. In view of (4.10) and (4.11), this can be done by directly estimating the vector $\boldsymbol{\beta}$, where its corresponding feasible region is a simple convex set in $\mathbb{R}^d$. First, define the *support matrix*

$$\boldsymbol{\Psi} = (\psi_{jk})_{d \times p}, \quad \psi_{jk} = \mathbb{I}\left((\{j\} \cap J_k) \neq \varnothing\right)$$

Then the moment conditions (4.5) can be expressed in matrix notation $\boldsymbol{\Psi}\boldsymbol{\beta} = \mathbf{1}$, where $\mathbf{1} = (1, 1, \ldots 1)^\top \in \mathbb{R}^d$. Similarly, for $\boldsymbol{\vartheta}$, Relation (4.26) can be expressed via

$$\boldsymbol{\vartheta} = \boldsymbol{\Lambda}\boldsymbol{\beta},$$

with $\boldsymbol{\Lambda} = \operatorname{supp}\left(\boldsymbol{\Psi}^\top \boldsymbol{\Psi}\right)$,where $\operatorname{supp}(\cdot)$ is the entry-wise signum function. Hence, a natural estimator for $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{b} \in B}{\arg\min} \|\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\Lambda}\boldsymbol{b}\|_2^2, \tag{4.28}$$

where B denotes the feasible region $B = \{\boldsymbol{b} \in \mathbb{R}_+^p : \boldsymbol{\Psi}\boldsymbol{b} = \mathbf{1}\}$.

We briefly comment on the large sample properties of $\hat{\boldsymbol{\beta}}$ and the optimization

problem (4.28).

Remark 4.9. For $\boldsymbol{\vartheta}^{-1} := (\vartheta_k^{-1})_{k=1}^p$, the CLT implies that $\sqrt{n}(\tilde{\boldsymbol{\vartheta}}^{-1} - \boldsymbol{\vartheta}^{-1}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$, where '$\xrightarrow{d}$' denotes convergence in distribution and the $p \times p$ matrix $\boldsymbol{\Sigma}$ has entries

$$\Sigma_{k\ell} = \mathrm{Cov}(\min_{j \in J_k} X_j^{-1}, \min_{j \in J_\ell} X_j^{-1}).$$

The delta method readily yields asymptotic normality for the naive, unconstrained estimators $\tilde{\boldsymbol{\vartheta}}$ and $\tilde{\boldsymbol{\beta}} = \boldsymbol{\Lambda}^{-1}\tilde{\boldsymbol{\vartheta}}$ in (4.27). For out estimator $\hat{\boldsymbol{\vartheta}} = \boldsymbol{\Lambda}\hat{\boldsymbol{\beta}}$, we have $\|\hat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\|_2 \leq \|\hat{\boldsymbol{\vartheta}} - \tilde{\boldsymbol{\vartheta}}\|_2 + \|\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\|_2 \leq 2\|\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}\|_2 \xrightarrow{p} 0$, where the second inequality holds due to the fact that $\hat{\boldsymbol{\beta}}$ is the unique minimizer in (4.28) and '$\xrightarrow{p}$' denotes convergence in probability. Consequently, $\hat{\boldsymbol{\beta}} = \boldsymbol{\Lambda}^{-1}\hat{\boldsymbol{\vartheta}}$ is a consistent estimator for $\boldsymbol{\beta}$. On the other-hand, the asymptotic distribution of $\hat{\boldsymbol{\beta}}$ is more delicate. In short, one should not expect asymptotic normality unless $\boldsymbol{\beta}$ lies within the interior of B. See Dupacova and Wets (1988) for a precise characterization of the asymptotic behavior of linear inequality constrained regression estimates with dependent errors.

Remark 4.10. One recognizes (4.28) as a standard convex quadratic program, which has a unique solution that can, theoretically, be obtained exactly using a finite number of steps. Many algorithms exist to compute the solution of quadratic programs but determining an efficient algorithm that exploits the unique structure of our problem would be desirable. For instance, the dimensionality of the parameter space $p = 2^d - (d + 1)$ can be massive even if the underlying dimension of the problem d is small. This seemingly high dimensional problem is countered by the fact that the feasible region B severely constrains the parameter space. The actual algorithm used for results in this paper is discussed in Section 4.7. Code in R implementing the TM model estimation is available from Yuen (2014).

The estimation procedure outlined above requires observations with marginals normalized to standard Fréchet (SMS). If the marginal parameters are known, one

can transform $\boldsymbol{V}$ to a SMS random vector $\boldsymbol{X}$ via (4.14). In practice, we observe $\boldsymbol{V}_1, \boldsymbol{V}_2, \ldots, \boldsymbol{V}_n$ with unknown marginal GEV parameters $\boldsymbol{\mu}, \boldsymbol{\sigma}$ and $\boldsymbol{\xi}$. As is commonly done with multivariate extremes, we first estimate the marginal GEV parameters and transform the margins to standard Fréchet before proceeding to fit the TM model. The exact method of estimating $\boldsymbol{\mu}, \boldsymbol{\sigma}$ and $\boldsymbol{\xi}$, while important, is not a focus of this work. Here we estimate the GEV parameters using maximum likelihood. Thus we summarize our estimation procedure as follows:

TM estimation procedure

1. Given $\boldsymbol{V}_1, \ldots, \boldsymbol{V}_n$ independent copies of a max-stable random vector $\boldsymbol{V}$ with

$$V_j \sim \text{GEV}(\mu_j, \sigma_j, \xi_j), j = 1, \ldots, d.$$

2. Marginally estimate $\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\xi}}$ via maximum likelihood.

3. Transform to Fréchet scale

$$\boldsymbol{X}_i = \boldsymbol{T}(\boldsymbol{V}_i; \hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\sigma}}, \hat{\boldsymbol{\xi}})$$

using equation (4.14).

4. For $\tilde{\boldsymbol{\vartheta}} = (\tilde{\vartheta}_1, \ldots, \tilde{\vartheta}_p)^\top$ set:

$$\tilde{\vartheta}_k^{-1} := \min\left\{1, \max\left\{|J_k|^{-1}, \frac{1}{n}\sum_{i=1}^{n} \min_{j \in J_k} X_{ij}^{-1}, \right\}\right\}$$

5. Solve the quadratic program

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{b} \in B}{\arg\min} \|\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\Lambda}\boldsymbol{b}\|_2^2, \; B = \{\boldsymbol{b} \in \mathbb{R}_+^p, \boldsymbol{\Psi}\boldsymbol{b} = \boldsymbol{1}\}.$$

To investigate the performance of our estimation procedure, we conduct a small simulation experiment using the max-linear models introduced in Section 4.2.1 with coefficient matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ defined in (4.24). The coefficients $\{\beta_{J_k}\}_{k=1}^p$ of the corresponding TM models are displayed as dark points '•' in Figures 4.3 and 4.4. For these simulations we specified homogeneous max-linear models with GEV parameters $\xi \in \{0.5, 1\}$, $\sigma = 1$ and $\mu = 1/\xi$. We carried out the TM estimation procedure

Table 4.2: Medians and inter-quartile range (IQR), in parentheses, for 1000 marginal GEV parameter estimates fitted to samples of size $n = 50$ generated from the max-linear model (4.23) with coefficient matrices $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ of (4.24). Reported are results for the first component V_1 only. Results for the remaining components are nearly identical.

	$\mu = 2$	$\sigma = 1$	$\xi = 0.5$
$\boldsymbol{A}_1$	1.990 (0.220)	0.977 (0.202)	0.516 (0.210)
$\boldsymbol{A}_2$	1.998 (0.228)	0.966 (0.226)	0.508 (0.205)
	$\mu = 1$	$\sigma = 1$	$\xi = 1$
$\boldsymbol{A}_1$	1.012 (0.248)	0.981 (0.321)	1.033 (0.309)
$\boldsymbol{A}_2$	1.009 (0.294)	0.988 (0.385)	1.039 (0.317)

described above using a sample size of $n = 50$. Note that we are estimating parameters of the TM model $\text{TM}(\boldsymbol{T}(\boldsymbol{V}; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi}))$, corresponding to the data generating max-stable random vector $\boldsymbol{V}$ which is well defined and identifiable. The estimating procedure was repeated for 1000 replications. A summary of the first component for the marginal GEV estimates are given in Table 4.2. GEV parameter estimates for components 2-5 are essentially identical and are omitted. Boxplots of the TM model coefficient estimates $\hat{\boldsymbol{\beta}}$ are shown in Figures 4.3 and 4.4. The estimates $\hat{\boldsymbol{\beta}}$ are accurate, with nearly all inter-quartile ranges (IQR's) concentrating around the true parameters. The variances of the estimates increase for the heavier tailed scenario $\xi = 1$. We conducted simulations with larger sample size $n \in \{100, 1000\}$ and results mirrored those for $n = 50$ with correspondingly smaller variances and thus are not shown.

4.4 An application to Industry portfolios

To illustrate the use of the TM model in practice, we now apply our methodology to extreme losses of industry portfolios. The data were obtained from French (2014)

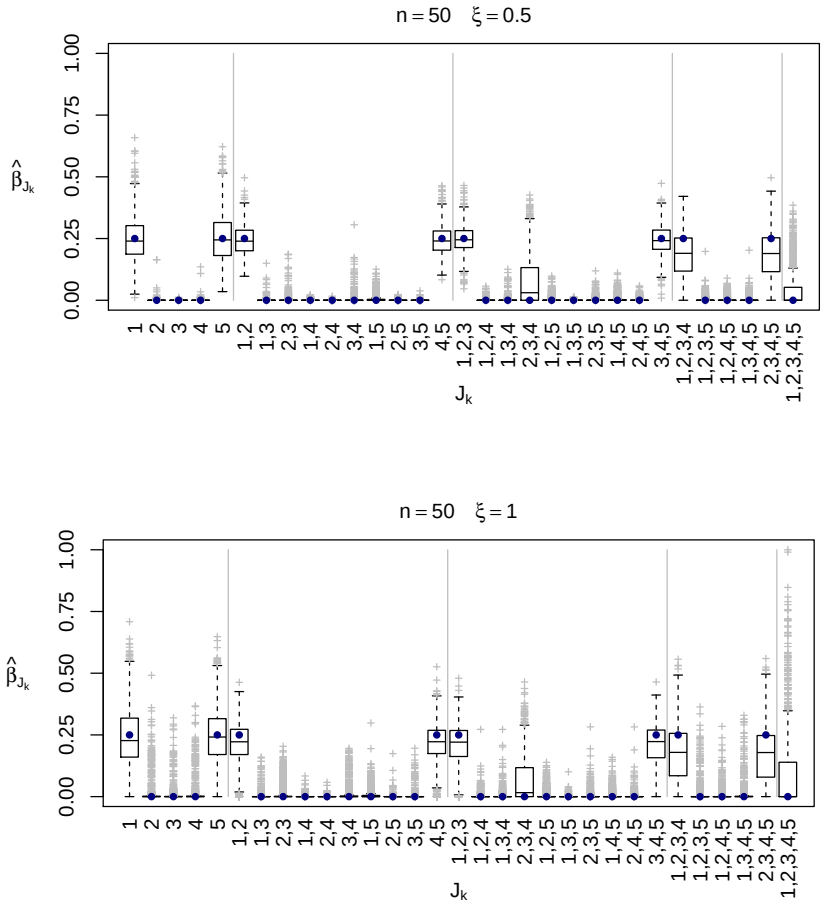


Figure 4.3: TM estimation results for max-linear model $\boldsymbol{A}_1$. Reported are boxplots of 1000 replicates of fitted TM model coefficients $\hat{\beta}_{J_k}$ using sample size $n = 50$. The true β_{J_k} for this model are indicated by the dark points •. The top panel corresponds to lighter tails ($\xi = 0.5$) vs heavier tails ($\xi = 1$) in the bottom panel.

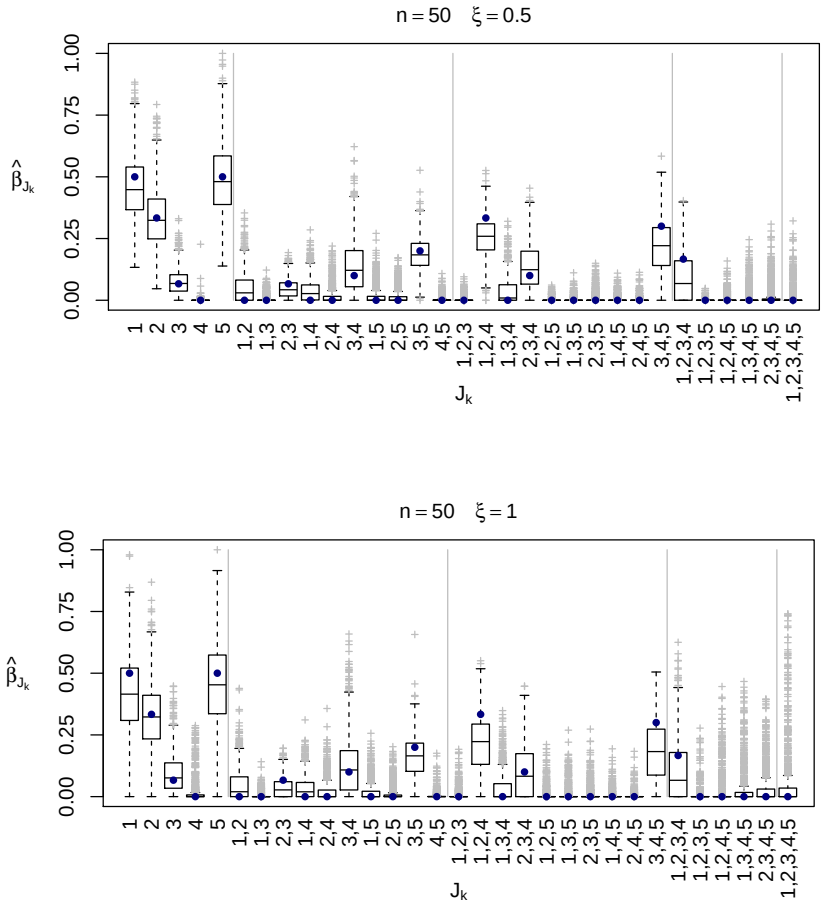


Figure 4.4: TM estimation results for max-linear model $\boldsymbol{A}_2$. Reported are boxplots of 1000 replicates of fitted TM model coefficients $\hat{\beta}_{J_k}$ using sample size $n = 50$. The true β_{J_k} for this model are indicated by the dark points •. The top panel corresponds to lighter tails ($\xi = 0.5$) vs heavier tails ($\xi = 1$) in the bottom panel.

and consist of daily returns from years (1927 - 2010) for five standardized industries: Consumer Goods, Manufacturing, Technology, Health Care, and Other – a final sector where all remaining assets are grouped together. The data have been pre-processed to have identical margins and thus are approximately homogeneous. Motivated by standard extreme value theory (see Section 4.9), we assume that the distribution of component-wise block maxima of negative daily returns is well approximated by a max-stable random vector $\boldsymbol{V}$. Here we select the yearly maximum negative daily return for each sector, resulting in a sample $\{\boldsymbol{V}_i\}_{i=1}^n$ of component-wise annual maxima for $n = 84$ years. In order to carry out our analysis, we must first re-normalize the data to standard Fréchet marginals. Typically one would fit marginal parameters μ_j, σ_j and ξ_j to each sector $j = 1, \ldots, d$. However, because the data are pre-processed, MLE estimates of GEV parameters based on individual sectors were nearly identical. Therefore we assume the marginal parameters are identical across sectors and we fit GEV parameters to the pooled data via maximum likelihood. This resulted in MLE estimates $\hat{\mu} = 2.59, \hat{\sigma} = 1.24$ and $\hat{\xi} = 0.39$. We then apply the transformation (4.14) to the data, $\boldsymbol{T}(\boldsymbol{V}_i; \hat{\mu}\mathbf{1}, \hat{\sigma}\mathbf{1}, \hat{\xi}\mathbf{1})$, so that the transformed data has approximately standard Fréchet marginals. Time series plots of these data on original scale and standard Fréchet scale are shown in Figure 4.5.

We then fit the TM model using the procedure outlined in Section 4.3. Figure 4.6(a) shows the estimated TM coefficients $(\hat{\beta}_{J_k})_{k=1}^p$. Recall the $\hat{\beta}_{J_k}$'s are ordered lexicographically starting with singleton sets $J_1 = \{1\}, \cdots, \{5\}$, and ending with $J_p = \{1, \cdots, 5\}$ – the coefficient involving all variables in the portfolio.

The coefficients corresponding to singleton sets indicate the levels of extremal *independence* between the industries. The right most coefficient, on the other hand, reflects the level of *complete dependence*. For example, consider Health Care (sector 4): roughly 1/3 of the extreme losses are attributed to shocks that are independent of the remaining sectors, while another 1/3 are attributed to shocks affecting the

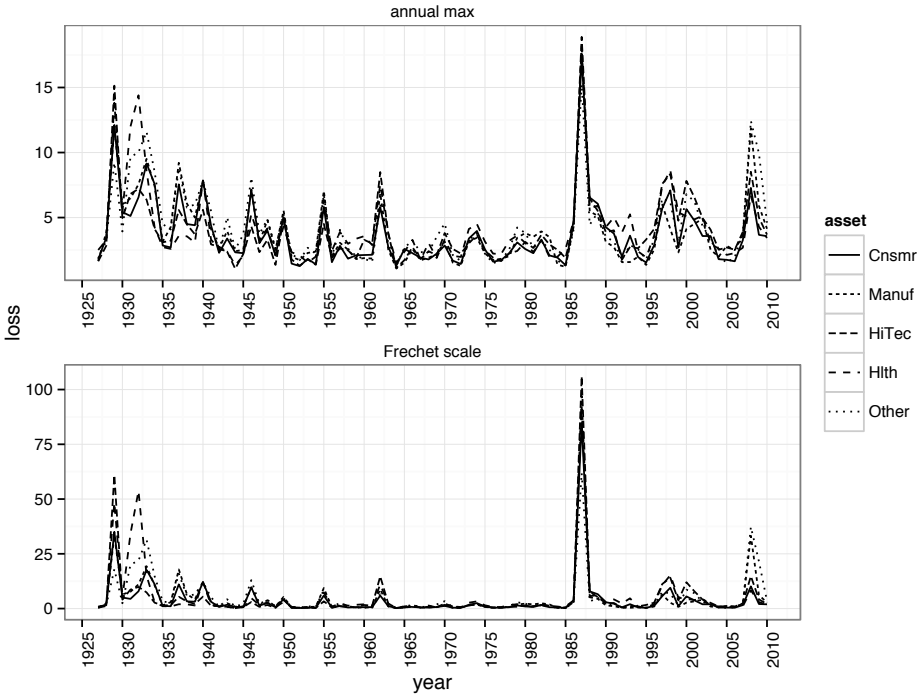
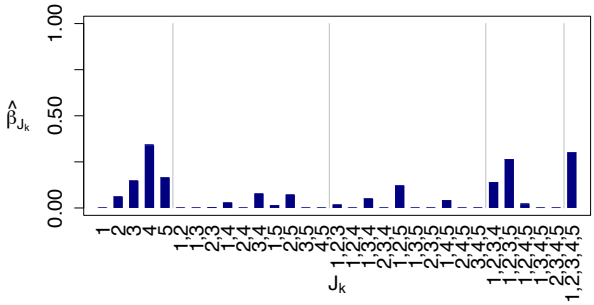
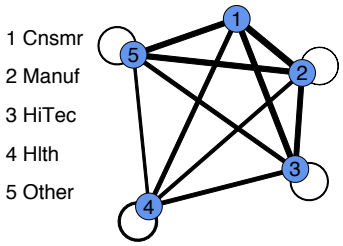


Figure 4.5: Asset-wise annual maximum of negative daily returns for 5 industry portfolio. **Top:** Original scale. **Bottom:** Fréchet scale.

entire market. This is further illustrated in Figure 4.6(b), where the edge connecting nodes i and j for example is weighted by $\hat{w}_{\boldsymbol{V}}(i,j) := \sum_{\{i,j\}\subset J_k} \hat{\boldsymbol{\beta}}_{J_k}$ which measures the proportion of time that losses affecting *both* variables i and j occur. Thus, for example, thinner lines indicate smaller degree of extremal dependence, as is the case for Health Care and Manufacturing sectors, while thicker lines indicate greater extremal dependence, e.g., between the Manufacturing and Consumer industries.



(a) Estimated coefficients of the TM model



(b) Tail dependence graph

	$\hat{w}(i,j)$				
	1	2	3	4	5
1	0.0000	0.8648	0.7732	0.5827	0.7649
2	-	0.0611	0.7267	0.4624	0.7782
3	-	-	0.1467	0.5681	0.5659
4	-	-	-	0.3409	0.3644
5	-	-	-	-	0.1648

(c) Estimated pairwise tail dependence coefficients

Figure 4.6: Illustration of the TM model applied to a 5-dimensional portfolio of stock market sectors.

4.4.1 Estimating upper bounds on VaR-max for portfolio losses

Motivated by Proposition 4.4, we aim to use the TM model to estimate upper bounds on VaR-max. We fit the TM model to monthly, quarterly and annual maxima, using the estimation procedure of Section 4.3. Recall that the data are pre-processed

to have identical margins and therefore we assume the homogeneous case where upper bounds hold with equality. Once the TM model is fitted, we estimate VaR-max directly using (4.17) with plug-in estimates

$$\mathrm{VaR}_\alpha(\boldsymbol{V}^\vee) \approx \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}} \left[\left(\frac{\hat{\vartheta}_{\mathrm{TM}(\boldsymbol{V})}(\{1,\ldots,d\})}{-\log \alpha} \right)^{\hat{\xi}} - 1 \right],$$

where $\hat{\mu}$,$\hat{\sigma}$, $\hat{\xi}$ are the marginal GEV paramter estimates from the TM model fit (see Section 4.3), and

$$\hat{\vartheta}_{\mathrm{TM}(\boldsymbol{V})}(\{1,\ldots,d\}) = \sum_{k=1}^{p} \hat{\beta}_{J_k}$$

To validate our methodology, we perform a standard backtesting procedure.

Backtesting procedure

1. Given block maxima $\{\boldsymbol{V}_i\}_{i=1}^n$
2. Set level $\alpha \in (0,1)$
3. Set length of training window m, where $m < n$.
4. For $i = 1$ to $n - m$
   (a) Fit TM model $\boldsymbol{W}_i$ to the training set $\{\boldsymbol{V}_k\}_{k=i}^{m+i-1}$
   (b) Calculate $\mathrm{VaR}_\alpha(\boldsymbol{W}_i^\vee)$ using formula (4.17)
   (c) Test: $E_i = \mathbf{1}_{\left[\boldsymbol{V}_{m+i}^\vee \leq \mathrm{VaR}_\alpha(\boldsymbol{W}_i^\vee)\right]}$
5. Compare level α to $\bar{E} = \frac{1}{n-m} \sum_{i=1}^{n-m} E_i$

Table 4.3 summarizes the results of the backtest where we considered $\alpha \in \{.900, .950, .975, .990\}$. The lengths of the training window m were chosen a priori to try to maximize the number of tests while maintaining a sufficient history,

and not posteriori for example, to minimize test error. For reference, we provide naive binomial standard errors that assume the test results $E_1, \ldots, E_{n-m}$ are independent. In reality, the errors should be larger. A longer time series would be required for significance at the α levels we consider. Nonetheless the results are encouraging. The quarterly VaR-max is well estimated while the monthly VaR-max appears to be slightly underestimated. This suggests, perhaps, that the block size corresponding to monthly time scales is not large enough to be approximated by a limiting max-stable model. While the annual VaR-max appears to be underestimated, the number of tests, 33, is far too small to make a significant conclusion. For visual reference, we include Figure 4.7 which traces the resulting estimated levels of VaR-max along with time series for annual maxima. Plots for monthly and quarterly block–maxima are very similar and thus omitted.

Remark 4.11. Our main goal in this study was not to estimate VaR–max. This can be done in a classical manner by modeling the one–dimensional time series of maximum portfolio losses and then applying established tools from univariate EVT (see e.g., Embrechts et al. (1997)). We compared this classical EVT approach to that of the TM model and results for VaR-max were virtually identical. The estimated TM model, however, gives additional insights to the extremal dependence structure of the portfolio. The above limited back–testing experiment coupled with validation by the classical EVT approach suggests that although high–dimensional, the estimates of the TM model are rather robust in practice.

4.5 Scaling VaR–max of block maxima to daily VaR–max

Let $\boldsymbol{R} = (R_i)_{i=1}^d$ be a random vector modeling negative daily returns (losses) of a d-asset portfolio. Suppose that (4.1) holds. The tail behavior of the maximum daily loss $\boldsymbol{R}^\vee := \max_{i=1,\cdots,d} R_i$ is governed by the assets with the heaviest tail indices

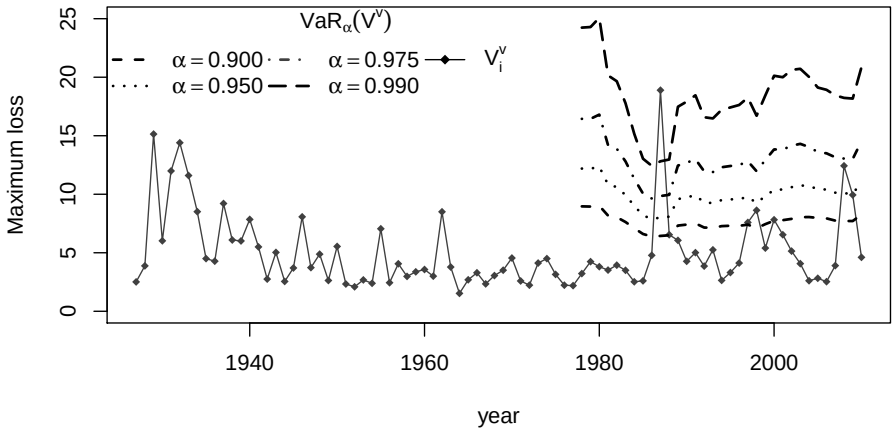


Figure 4.7: Backtesting time series for annual block maxima. Solid line indicates maxima $\{\boldsymbol{V}_i^\vee\}_{i=1}^n$ of the 5 industry portfolio in Figure 4.5. Broken lines correspond to the series of estimated $\mathrm{VaR}_\alpha(\boldsymbol{V}^\vee)$ at $\alpha \in \{.900, .950, .975, .990\}$ used in the backtest.

$\xi_i = \xi^* := \max_{i=1,\cdots,d} \xi_j$, where ξ_j stands for the shape parameter of the GEV marginal distribution of the max–stable vector $\boldsymbol{V} = (V_i)_{i=1}^d$. For simplicity, in order to illustrate how VaR–max scales, we shall suppose that the tail indices $\xi^* = \xi_i,\ i = 1, \cdots, d$ of all d variables are *equal*. Otherwise, one should focus on the sub–vector of heaviest tails. We shall suppose moreover that $\xi^* > 0$ and

$$a_{n,i} := n^{\xi^*} \quad \text{and} \quad b_{n,i} = 0,\ i = 1, \cdots, n. \tag{4.29}$$

That is, we are in the simple situation with trivial slowly varying function and common heavy–tailed marginals. In this case we have the following simple result. Its proof is given in Section 4.9.

Lemma 4.12. *Relations* (4.1) *and* (4.29) *imply*

$$VaR_\alpha(\boldsymbol{R}^\vee) \sim VaR_\alpha(\boldsymbol{V}^\vee), \quad as\ \alpha \uparrow 1, \tag{4.30}$$

Table 4.3: TM model backtesting results for 5 industry portfolio. Empirical coverage rates ($\hat{\alpha}$) and naive binomial standard errors for Monthly, Quarterly and Annual TM bounds on VaR-max. The training length and number of tests are given by m and $(n-m)$ respectively.

Block:	Monthly	Quarterly	Annual
m $(n-m)$:	60 (947)	100 (235)	50 (33)
α	$\hat{\alpha}$ (Binomial s.e.)		
0.900	0.873 (0.010)	0.872 (0.020)	0.818 (0.052)
0.950	0.936 (0.007)	0.945 (0.014)	0.939 (0.038)
0.975	0.963 (0.005)	0.975 (0.010)	0.970 (0.027)
0.990	0.982 (0.003)	0.996 (0.007)	0.970 (0.017)

where $\boldsymbol{V}^{\vee} := \max_{i=1,\cdots,d} V_i$.

Consider now a period of M consecutive days and focus on the maximum loss over all assets in this period. We wish to determine how quantiles of this maximum M-day loss scale as a function of M. By (4.1),

$$\frac{1}{M^{\xi^*}} \max_{k=1,\cdots,M} \boldsymbol{R}_k^{\vee} \xrightarrow{d} \boldsymbol{V}^{\vee} := \max_{i=1,\cdots,d} V_i, \quad \text{as } M \to \infty.$$

This suggests that for relatively large M, we have

$$\mathrm{VaR}_{\alpha}\Big(\max_{k=1,\cdots,M} \boldsymbol{R}_k^{\vee} \Big) \approx M^{\xi^*} \mathrm{VaR}_{\alpha}(\boldsymbol{V}^{\vee})$$

(see e.g., Proposition 0.1 in Resnick, 1987). On the other hand, Lemma 4.12 can be applied in the case when $\boldsymbol{R}^{\vee}$ and $\boldsymbol{V}^{\vee}$ are replaced with $\max_{k=1,\cdots,M} \boldsymbol{R}_k^{\vee}$ and $M^{\xi^*}\boldsymbol{V}^{\vee}$, respectively. Therefore, for fixed M, we obtain

$$\mathrm{VaR}_{\alpha}\Big(\max_{k=1,\cdots,M} \boldsymbol{R}_k^{\vee} \Big) \sim M^{\xi^*} \mathrm{VaR}_{\alpha}(\boldsymbol{V}^{\vee}), \quad \text{as } \alpha \uparrow 1.$$

This and Relation (4.30) suggest the following first order scaling correction for VaR–max:

$$\mathrm{VaR}_\alpha(\boldsymbol{R}^\vee) \approx M^{-\xi^*}\mathrm{VaR}_\alpha\Big(\max_{k=1,\cdots,M}\boldsymbol{R}_k^\vee\Big),$$

where for fixed M the above approximation becomes an asymptotic equivalence, as $\alpha \uparrow 0$.

In reality, however, the returns $\boldsymbol{R}_k,\ k = 1, 2, \cdots$ are *dependent* in k. Suppose that this time series is stationary and has homogeneous heavy–tailed marginals. Under very general conditions, the degree of the temporal dependence of the maximum loss time series $\{\boldsymbol{R}_k^\vee\}$ can then be quantified in terms of the *extremal index* parameter $\theta \in (0, 1]$ as follows:

$$\frac{1}{M^{\xi^*}}\max_{k=1,\cdots,M}\boldsymbol{R}_k^\vee \xrightarrow{d} \theta^{\xi^*}\boldsymbol{V}^\vee, \quad \text{as } n \to \infty.$$

(see e.g., Leadbetter et al., 1983). This shows that the scaling of VaR–max should be suitably adjusted to account for dependence as follows:

$$\mathrm{VaR}_\alpha(\boldsymbol{R}^\vee) \approx (\theta M)^{-\xi^*}\mathrm{VaR}_\alpha\Big(\max_{k=1,\cdots,M}\boldsymbol{R}_k^\vee\Big) \approx (\theta M)^{-\xi^*}\mathrm{VaR}_\alpha(\boldsymbol{V}^\vee). \tag{4.31}$$

In practice, we recommend estimating the extremal index θ and applying (4.31) with ξ^* equal to the maximum shape parameter of the fitted GEV marginal distributions. Numerous procedures for the estimation of θ exist in the literature, see e.g., Hamidieh et al. (2009) and the references therein. This discussion suggests the following methodology for *scaling of VaR–max*.

Scaling of VaR–max

1. Given a time series $\{\boldsymbol{R}_k\}_{k=1}^n$, estimate the marginal GEV parameters and set $\xi^* = \max_{i=1,\cdots,d} \xi_i$.

2. Estimate the extremal index θ of the time series $\{\boldsymbol{R}_k^\vee\}_{k=1}^n$ of maximum portfolio losses.

3. Consider block–maxima $\boldsymbol{V}_i^{(M)} := \max_{k=1,\cdots,M} \boldsymbol{R}_{k+M(i-1)}$, $i = 1, \cdots, [n/M]$ over non–overlapping blocks of M consecutive $\boldsymbol{R}_k$'s.

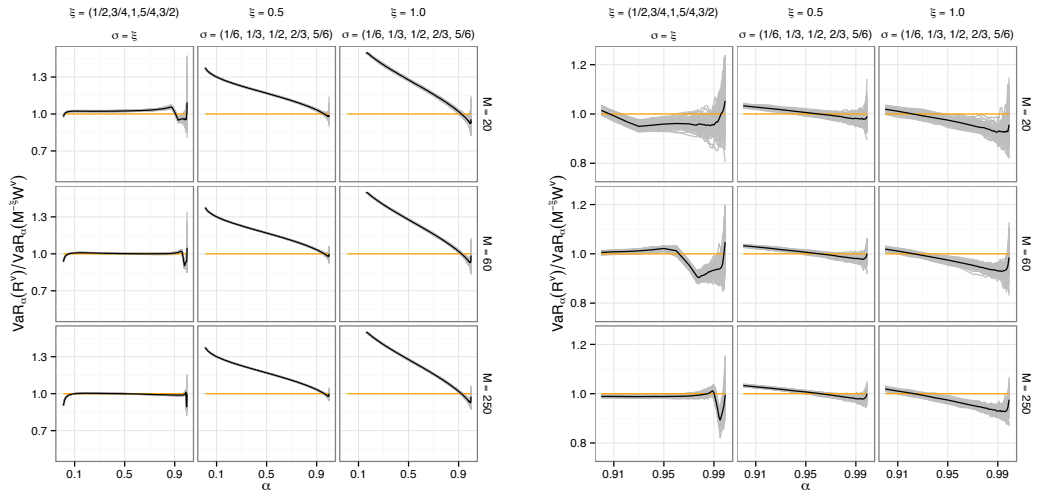
4. Assuming the $\boldsymbol{V}_i^{(M)}$'s are max–stable, estimate the parameters of the TM model $\boldsymbol{W} = \mathrm{TM}(\boldsymbol{V}_1^{(M)})$.

5. Given $\alpha \in (0,1)$, compute the rescaled VaR–max as in (4.31), where $\mathrm{VaR}_\alpha(\boldsymbol{V}^\vee)$ is replaced by $\mathrm{VaR}_\alpha(\boldsymbol{W}^\vee)$.

To investigate the scaling of VaR–max in a more realistic setting, we considered the linear model

$$\widetilde{\boldsymbol{R}} = \boldsymbol{A}\boldsymbol{Z} + \boldsymbol{\epsilon},$$

where $\boldsymbol{Z} = (Z_j)_{j=1}^p$ is a vector of iid standard Fréchet variables and $\boldsymbol{\epsilon} = (\epsilon_i)_{i=1}^d \sim \mathcal{N}(\boldsymbol{0}, \sigma_\epsilon^2 I)$ is Gaussian noise. The $(d \times p)$ matrix $\boldsymbol{A} = (a_{i,j})$ is assumed to have non–negative entries. In this case, it is well known that $\widetilde{\boldsymbol{R}}$ belongs to the maximum domain of attraction of the max–linear model in (4.21). To study the effect of varying marginal distributions on the gap in the stochastic upper bounds, we also transformed the marginal distributions to GEV with various scales and shape parameters of Table 4.1 and consider returns $\boldsymbol{R} := \boldsymbol{T}(\widetilde{\boldsymbol{R}}; \boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\xi})$, with $\boldsymbol{T}$ as in (4.14). We focused on the two cases $\boldsymbol{A} = \boldsymbol{A}_1$ and $\boldsymbol{A} = \boldsymbol{A}_2$ given in (4.24).



(a) Linear model with $\boldsymbol{A}_2$

(b) Linear model with $\boldsymbol{A}_2$ zoomed to $\alpha \in (.9, .999)$

Figure 4.8: VaR-max ratios for daily maxima $\boldsymbol{R}^\vee$ versus scaled TM model $M^{-\xi^*}\boldsymbol{W}^\vee$ using block size $M \in \{20, 60, 250\}$. Grey lines indicate 100 empirical versions of the ratio $\mathrm{VaR}_\alpha(\boldsymbol{R}^\vee)/\mathrm{VaR}_\alpha(M^{-\xi^*}\boldsymbol{W}^\vee)$ and the dark line reports the median. Columns indicate different values of the shape and scale parameters for the margins of $\boldsymbol{W}$.

Figure 4.8 illustrates the accuracy of the proposed scaling methodology in the simple case of temporally independent returns, which are governed by a linear factor model with non–homogeneous marginals. Observe that the TM-based upper bound fails for a wide range of levels α. As expected, however, the bound starts to work reasonably well for the range of large values of $\alpha \in (0.9, 0.999)$, which is of most interest in applications.

4.6 Discussion

In summary, we have presented a theoretically sound methodology for determining upper bounds on Value-at-Risk at high levels of the maximum loss within a portfolio (VaR-max). The bounds are achieved via the lower orthant dominance property of the Tawn-Molchanov (TM) model. We developed a simple regression based procedure

for fitting the TM model which, as a by-product, yields simple and interpretable summary statistics of tail dependence. The finite sample behavior of the theoretical Tawn-Molchanov bounds for VaR-max was illustrated using numerical simulations and a practical example with industry portfolios.

4.7 Notes on the quadratic program

Here we provide more precise details on the TM model estimation introduced in Section 4.3. Recall the quadratic program used to fit the TM model:

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{b}\in B}{\arg\min} \|\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\Lambda}\boldsymbol{b}\|_2^2, \quad B = \{\boldsymbol{b} \in \mathbb{R}_+^p : \boldsymbol{\Psi}\boldsymbol{b} = \mathbf{1}\}. \tag{4.32}$$

The equality constraints appearing in the feasible region B show that the solution to (4.32) is only determined by $p - d = 2^d - (d + 1)$ free parameters. Here, we provide an expression for the loss function in terms of the free parameters. Consider a partition of the coefficient vector $\boldsymbol{\beta} = (\ \boldsymbol{\beta}_{1:d} \quad \boldsymbol{\beta}_{\boldsymbol{P}}\ )^\top$ and support matrix

$$\boldsymbol{\Psi}_{d\times 2^d-1} = (\ I_{d\times d} \quad \boldsymbol{P}\ ). \tag{4.33}$$

Given the moment constraints $\boldsymbol{\Psi}\boldsymbol{\beta} = \mathbf{1}$ appearing in B, $\boldsymbol{\beta}_{1:d} = \mathbf{1} - \boldsymbol{P}\boldsymbol{\beta}_P$. Now, with respect to the free parameters $\boldsymbol{\beta}_P$, the feasible region B in (4.32) translates into:

$$B_{\boldsymbol{P}} = \{\boldsymbol{b} \in \mathbb{R}_+^{p-d} : \boldsymbol{P}\boldsymbol{b} \leq \mathbf{1}\}.$$

Note that $B_{\boldsymbol{P}}$ is a compact convex set in $\mathbb{R}^{p-d}$ with non–empty interior. By the lexicographical ordering of $J_1 = \{1\},\ J_2 = \{2\}, \ldots, J_d = \{1, \ldots, d\}$, we have $\vartheta_k = 1$ for $k = 1, \ldots, d$. Hence the first d elements of the vector $\tilde{\boldsymbol{\vartheta}} = (\ \mathbf{1} \quad \tilde{\boldsymbol{\vartheta}}_{\boldsymbol{P}}\ )^\top$ are set to

unity. We will verify that

$$\min_{\boldsymbol{\beta}\in B} \|\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\Lambda}\boldsymbol{\beta}\|_2^2 = \min_{\boldsymbol{\beta}_{\boldsymbol{P}}\in B_{\boldsymbol{P}}} \|(\tilde{\boldsymbol{\vartheta}}_{\boldsymbol{P}} - \boldsymbol{P}^\top \mathbf{1}) - \boldsymbol{\Lambda}_{\boldsymbol{P}}\boldsymbol{\beta}_{\boldsymbol{P}}\|_2^2 \tag{4.34}$$

where $\boldsymbol{\Lambda}_{\boldsymbol{P}} = \operatorname{supp}\left(\boldsymbol{P}^\top \boldsymbol{P}\right) - \boldsymbol{P}^\top \boldsymbol{P}$. Recall

$$\boldsymbol{\Lambda} = \operatorname{supp}\left(\boldsymbol{\Psi}^\top \boldsymbol{\Psi}\right) = \begin{pmatrix} I & \boldsymbol{P} \\ \boldsymbol{P}^\top & \operatorname{supp}\left(\boldsymbol{P}^\top \boldsymbol{P}\right) \end{pmatrix}.$$

Therefore

$$\boldsymbol{\Lambda}\boldsymbol{\beta} = \begin{pmatrix} I & \boldsymbol{P} \\ \boldsymbol{P}^\top & \operatorname{supp}\left(\boldsymbol{P}^\top \boldsymbol{P}\right) \end{pmatrix} \begin{pmatrix} \mathbf{1} - \boldsymbol{P}\boldsymbol{\beta}_{\boldsymbol{P}} \\ \boldsymbol{\beta}_{\boldsymbol{P}} \end{pmatrix} = \begin{pmatrix} \mathbf{1} \\ \boldsymbol{P}^\top \mathbf{1} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\Lambda}_{\boldsymbol{P}}\boldsymbol{\beta}_{\boldsymbol{P}} \end{pmatrix}.$$

and it follows that

$$\tilde{\boldsymbol{\vartheta}} - \boldsymbol{\Lambda}\boldsymbol{\beta} = \begin{pmatrix} \mathbf{1} \\ \tilde{\boldsymbol{\vartheta}}_{\boldsymbol{P}} \end{pmatrix} - \begin{pmatrix} \mathbf{1} \\ \boldsymbol{P}^\top \mathbf{1} \end{pmatrix} - \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\Lambda}_{\boldsymbol{P}}\boldsymbol{\beta}_{\boldsymbol{P}} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \left(\tilde{\boldsymbol{\vartheta}}_{\boldsymbol{P}} - \boldsymbol{P}^\top \mathbf{1}\right) - \boldsymbol{\Lambda}_{\boldsymbol{P}}\boldsymbol{\beta}_{\boldsymbol{P}} \end{pmatrix}.$$

Hence, we carry out optimization with the criterion (4.34).

For results in this paper we solve (4.34) using the quadratic program solver `solve.QP` in the `R` package `quadprog` (Turlach and Weingessel, 2013). This is likely inefficient since we do not take advantage of the unique structure of the problem. Given that moderate dimensions d lead to a very high dimensional parameter $2^d - (d+1)$, a solution to (4.34) in the high dimensional case, say $d > 16$, requires a specialized implementation of this quadratic program. Some key observations include the following

- The matrix $\boldsymbol{\Lambda}_{\boldsymbol{P}}$ is full rank, dense, yet highly structured and fully determined by the dimension d. Consequently, the loss function can be calculated without storing $\boldsymbol{\Lambda}_{\boldsymbol{P}}$ and thus requiring relatively little memory with respect to the

ambient dimension $2^d - (d+1)$.

- Similar properties hold for the vector $\tilde{\boldsymbol{\vartheta}}$ which can be computed in parallel and is fully determined by the data, where the dimension $d \times n$ can be handled easily for many applications.

- The problem may be cast into a penalized regression framework where the feasible region $B_{\boldsymbol{P}} = \{\boldsymbol{b} \in \mathbb{R}_+^{p-d} : \boldsymbol{P}\boldsymbol{b} \leq \mathbf{1}\}$ corresponds to d L_1 penalties plus the positivity penalty

$$\phi(\boldsymbol{\beta}_{\boldsymbol{P}}) = \begin{cases} 0 & \text{if } \min_k \beta_{J_k} \geq 0 \\ \infty & \text{otherwise} \end{cases}.$$

4.8 Notes on the TM model and k-way extremal dependence

Let $\boldsymbol{X}$ be an SMS random vector in $\mathbb{R}^d$. The coefficients w_J in (4.12) associated with the Tawn–Molchanov model TM($\boldsymbol{X}$) can be related to higher order tail dependence coefficients of $\boldsymbol{X}$. Indeed, following for example p. 259 of de Haan and Ferreira (2006) (see also the quantity $R(1,1)$ on page 225 therein) introduce

$$\lambda_{\boldsymbol{X}}(i_1, \cdots, i_k) := \lim_{x \to \infty} x\mathbb{P}\left(\frac{1}{1 - G_i(X_i)} > x,\ i = i_1, \cdots, i_k\right). \tag{4.35}$$

The $\lambda_{\boldsymbol{X}}(i_1, \cdots, i_k)$'s may be viewed as k-way tail dependence coefficients.

Proposition 4.13. *Let $\boldsymbol{R}$ be in the max domain of attraction of the max–stable vector $\boldsymbol{X}$. Then, for all $J := \{i_1, \cdots, i_k\} \subset \{1, \cdots, d\}$, we have:*

$$\lambda_{\boldsymbol{X}}(i_1, \cdots, i_k) = \lim_{x \to \infty} x\mathbb{P}\left(\frac{1}{1 - F_i(R_i)} > x,\ i = i_1, \cdots, i_k\right) \tag{4.36}$$

and moreover $\lambda_{\boldsymbol{X}}(i_1,\cdots,i_k)=\lambda_{\mathrm{TM}(\boldsymbol{X})}(i_1,\cdots,i_k)=w_{\boldsymbol{X}}(i_1,\cdots,i_k)$, *where* $w_{\boldsymbol{X}}$ *is as in* (4.12).

Proof. Let $A_i=\{\boldsymbol{z}=(z_j)_{j=1}^d\in\mathbb{R}_+^d\,:\,z_i>1\},\ i=1,\cdots,d$ and set $A:=\cap_{i\in J}A_i$. Thus, the right–hand sides of (4.35) and (4.36) equal $\lim_{x\to\infty}x\mathbb{P}((1-\boldsymbol{G}(\boldsymbol{X}))^{-1}\in xA)$ and $\lim_{x\to\infty}x\mathbb{P}((1-\boldsymbol{F}(\boldsymbol{V}))^{-1}\in xA)$, respectively. Using the homogeneity of the exponent measure ν one can show that $A=\cap_{i\in J}A_i$ is a continuity set of ν. Thus, the multivariate regular variation property (4.3) implies that the limits in (4.35) and (4.36) exist and are both equal to $\nu_{\boldsymbol{X}}(A)\equiv\nu_{\boldsymbol{X}}(\cap_{i\in J}A_i)$. To complete the proof, it remains to show that

$$\nu_{\boldsymbol{X}}(A)=\nu_{\mathrm{TM}(\boldsymbol{X})}(A)=w_{\boldsymbol{X}}(J). \tag{4.37}$$

The CDF of the max–stable random vector $\boldsymbol{X}$, expressed through its exponent measure is:

$$\mathbb{P}(\boldsymbol{X}\le\boldsymbol{x})=\exp\Big\{-\nu_{\boldsymbol{X}}([\boldsymbol{0},\boldsymbol{x}]^c)\Big\},\quad \boldsymbol{x}\in\mathbb{R}_+^d \tag{4.38}$$

(see e.g., Ch. 5 in Resnick (1987)). This, using the definition of the extremal coefficients, yields

$$\vartheta_{\boldsymbol{X}}(K)=\nu_{\boldsymbol{X}}(\cup_{i\in K}A_i),\quad \text{for all } \emptyset\neq K\subset\{1,\cdots,d\}.$$

Since $\vartheta_{\boldsymbol{X}}(K)=\vartheta_{\mathrm{TM}(\boldsymbol{X})}(K)$, for all $K\subset\{1,\cdots,d\}$, we therefore have $\nu_{\boldsymbol{X}}(\cup_{i\in K}A_i)=\nu_{\mathrm{TM}(\boldsymbol{X})}(\cup_{i\in K}A_i)$. That is, the exponent measures of $\boldsymbol{X}$ and $\mathrm{TM}(\boldsymbol{X})$ coincide over arbitrary unions of the sets $A_i,\ i=1,\cdots,d$. Therefore, using the inclusion–exclusion formula, one obtains also that

$$\nu_{\boldsymbol{X}}(\cap_{i\in K}A_i)=\nu_{\mathrm{TM}(\boldsymbol{X})}(\cap_{i\in K}A_i),\quad \text{for all } \emptyset\neq K\subset\{1,\cdots,d\}.$$

This proves the first equality in (4.37). To establish the second one, note that by

(4.4) and (4.38), the exponent measure $\nu_{\mathrm{TM}(\boldsymbol{X})}$ has the following form:

$$\nu_{\mathrm{TM}(\boldsymbol{X})}(B) = \sum_{k=1}^{p} \beta_{J_k} \nu_1(\{x > 0 \,:\, x\mathbf{1}_{J_k} \in B\}),$$

where $\mathbf{1}_{J_k} = (1_{J_k}(i))_{i=1}^{d} \in \mathbb{R}_+^d$ and ν_1 is the measure on $(0,\infty)$ for which $\nu_1(x,\infty) = x^{-1},\ x > 0$. Using this representation, it can be seen directly that

$$\nu_{\mathrm{TM}(\boldsymbol{X})}(\cap_{i\in J} A_i) = \sum_{k=1}^{p} \beta_{J_k} \nu_1(\{x > 0 \,:\, x\mathbf{1}_{J_k} \in \cap_{i\in J} A_i\}) = \sum_{k\,:\ J\subset J_k} \beta_{J_k},$$

because the term involving the measure ν_1 is either zero or one, and it equals one, only if $1_J \leq 1_{J_k}$. This yields the second equality in (4.37). □

Remark 4.14. By analogy with the bivariate case, one can visualize the *entire* tail dependence structure of the TM model through a weighted hyper–graph, where the hyper–edge corresponding to a set $J \subset \{1, \cdots, d\}$ is weighted by $w_{\boldsymbol{X}}(J)$.

4.9 Proofs

The cumulative distribution function of any SMS random vector is completely determined by either one of the following two characterizations.

- *(spectral measure)* The unique finite measure H on $\mathbb{S}_+^{d-1} = \{\boldsymbol{w} \in \mathbb{R}_+^d : \|\boldsymbol{w}\|_1 = 1\}$ such that

$$\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) = \exp\left[-\int_{\mathbb{S}_+^{d-1}} \bigvee_{j=1}^{d} \frac{w_j}{x_j} H(d\boldsymbol{w})\right]. \tag{4.39}$$

  To ensure standard Fréchet margins, H must satisfy the moment conditions $\int_{\mathbb{S}_+^{d-1}} w_j H(d\boldsymbol{w}) = 1,\ j = 1, \ldots, d$.

- *(dependency set)* (see Molchanov, 2008) The largest compact convex set $\Delta \subset \mathbb{R}_+^d$

such that for all $\boldsymbol{x} \in \mathbb{R}_+^d$

$$-\log \mathbb{P}\left(X_j \leq 1/x_j, j = 1, \ldots, d\right) = \sup\left\{\boldsymbol{x}^\top \boldsymbol{y} : \boldsymbol{y} \in \Delta\right\}. \tag{4.40}$$

Here, the moment conditions translate into $\sup\{y_j : \boldsymbol{y} \in \Delta\} = 1,\ j = 1, \ldots, d$.

The next result shows that nesting of dependency sets between SMS random vectors implies stochastic dominance in the lower orthant order.

Proposition 4.15 (See also bottom of p. 242 in Molchanov, 2008)**.** *Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be two SMS random vectors in $\mathbb{R}_+^d$ having respective dependency sets $\Delta_{\boldsymbol{X}}$ and $\Delta_{\boldsymbol{Y}}$. If $\Delta_{\boldsymbol{X}} \subset \Delta_{\boldsymbol{Y}}$ then*

$$\mathbb{P}(\boldsymbol{X} \leq \boldsymbol{x}) \geq \mathbb{P}(\boldsymbol{Y} \leq \boldsymbol{x}).$$

Proof. In view of (4.40), the nesting $\Delta_{\boldsymbol{X}} \subset \Delta_{\boldsymbol{Y}}$ implies that for all $\boldsymbol{z} \in \mathbb{R}_+^d$,

$$-\log \mathbb{P}\left(X_j \leq 1/z_j, j = 1, \ldots, d\right) \leq -\log \mathbb{P}\left(Y_j \leq 1/z_j, j = 1, \ldots, d\right).$$

This completes the proof. □

Proof of Proposition 4.2. Let $\boldsymbol{X}$ be an arbitrary SMS random vector. Let $\Delta_{\boldsymbol{X}}$ and $\Delta_{\mathrm{TM}(\boldsymbol{X})}$ be the dependency sets of $\boldsymbol{X}$ and $\mathrm{TM}(\boldsymbol{X})$ respectively. Then, by Theorem 32 of Strokorb and Schlather (2013), we have $\Delta_{\boldsymbol{X}} \subset \Delta_{\mathrm{TM}(\boldsymbol{X})}$ and the lower orthant dominance follows immediately from Proposition 4.15. □

Proof of Lemma 4.12. By (4.3), we have that

$$\mathbb{P}(\boldsymbol{R}^\vee > x) \sim \mathbb{P}(\boldsymbol{V}^\vee > x), \quad \text{as } x \to \infty. \tag{4.41}$$

Let $f_R(x) := 1/\mathbb{P}(\boldsymbol{R}^\vee > x),\ x \geq 0$ and observe that f_R is monotone non–decreasing and regularly varying at infinity with exponent $1/\xi^*$. Thus, by Theorem 1.5.12 in

Bingham et al. (1987), there exists unique (up to asymptotic equivalence) asymptotic inverse function g_R, which is regularly varying with exponent ξ^* and such that

$$f_R(g_R(x)) \sim g_R(f_R(x)) \sim x, \quad \text{as } x \to \infty. \tag{4.42}$$

In fact, $g_R(x) = f_R^{\leftarrow}(x) := \inf\{y \geq 0 : f(y) > x\}$ is such an inverse. In view of Definition 4.3,

$$\mathrm{VaR}_\alpha(\boldsymbol{R}^\vee) = f_R^{\leftarrow}(1/(1-\alpha)). \tag{4.43}$$

Similarly, by letting $f_V(x) := 1/\mathbb{P}(\boldsymbol{V}^\vee > x)$, we obtain that $f_V(g_V(x)) \sim g_V(f_V(x)) \sim x$, as $x \to \infty$ and

$$\mathrm{VaR}_\alpha(\boldsymbol{V}^\vee) = f_V^{\leftarrow}(1/(1-\alpha)) \equiv g_V(1/(1-\alpha)). \tag{4.44}$$

Note, however, that by (4.41), we have $f_R(x) \sim f_V(x),\ x \to \infty$. This fact and the regular variation property of g_R imply that

$$f_V(g_R(x)) \sim g_R(f_V(x)) \sim x, \quad \text{as } x \to \infty.$$

That is, $g_R \equiv f_R^{\leftarrow}$ and $g_V \equiv f_V^{\leftarrow}$ are both asymptotic inverses of f_V. The uniqueness part in Theorem 1.5.12 in Bingham et al. (1987) shows that $f_R^{\leftarrow}(x) \sim f_V^{\leftarrow}(x)$, as $x \to \infty$, which by (4.43) and (4.44) yields (4.30). □

CHAPTER 5

Bounding Value-at-Risk for the sum of dependent losses

5.1 Introduction

As revealed in the previous chapter, Value-at-Risk (VaR) is one of the predominant risk measures used in determining minimum capital requirements placed upon financial institutions in order to cover potential losses in the market. In essence, VaR is the largest loss having a 'reasonable chance' of occurring though the placement of a risky bet. Formally, if the random variable $X \in \mathbb{R}$ represents a loss (negative return) on an asset after a fixed holding period, and $(1-\alpha) \in (0,1)$ corresponds to a small probability representing 'reasonable chance', then recall

Definition 5.1. The Value-at-Risk of a random variable $X \in \mathbb{R}$ at the level $\alpha \in (0,1)$, denoted $\mathrm{VaR}_\alpha(X)$ is defined as

$$\mathrm{VaR}_\alpha(X) := \inf\{x \in \mathbb{R} | \mathbb{P}(X \leq x) \geq \alpha\}.$$

In practice, financial institutions deal with a portfolio $\{1, \ldots, d\}$ of dependent losses $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top \in \mathbb{R}^d$, in which case capital requirements are determined by the Value-at-Risk for the sum of losses $\mathrm{VaR}_\alpha(S)$, where $S := X_1 + X_2 +$

$\cdots + X_d$. In these scenarios it is essential to account for tail dependence in the components of $\boldsymbol{X}$ due to the possibly *super-additive* nature of VaR, i.e. cases where

$$\sum_{j=1}^{d} \mathrm{VaR}_\alpha(X_j) < \mathrm{VaR}_\alpha(X_1 + \cdots + X_d).$$

Regulatory guidelines such as the Basel III (Bank for International Settlements, 2011) typically prescribe $\alpha \geq .99$. Hence, scenarios where α is close to the value 1 is of primary interest. Multivariate regular variation is a natural framework for characterizing $\mathrm{VaR}_\alpha(S)$ when $\alpha \approx 1$. To further illustrate, recall the following condition

Condition 5.2 (Multivariate Regular Variation, cf Theorem 6.1 of Resnick, 2007)**.** Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)^\top$ be a non-negative random vector with identical margins and $S = X_1 + X_2 + \cdots + X_d$. The vector $\boldsymbol{X} \in \mathbb{R}_+^d$ is multivariate regularly varying with index $-1/\xi$ if there exists

(i) A function $h(t)$ with $\lim_{t\to\infty} h(t) = \infty$ such that

$$\lim_{t\to\infty} t\mathbb{P}(X_1 > sh(t)) = s^{-1/\xi}, \text{ for all } s > 0,$$

(ii) a measure H on $\mathbb{S}_+^{d-1} = \{\boldsymbol{u} \in \mathbb{R}_+^d : u_1 + u_2 + \cdots + u_d = 1\}$ with finite mass $H(\mathbb{S}_+^{d-1}) = d$, called the *spectral measure*,

(iii) a scalar $\rho > 0$, possibly dependent on ξ and H,

such that for all $s > 0$

$$\lim_{t\to\infty} t\mathbb{P}\left(S > sh(t), \frac{\boldsymbol{X}}{S} \in A\right) = \rho s^{-1/\xi} H(A)/d, \tag{5.1}$$

for any $A \subset \mathbb{S}_+^{d-1}$ that is a continuity set of H.

If Condition 5.2 holds then we say $\boldsymbol{X} \in \mathrm{MRV}_+^d(-1/\xi)$ and it readily follows that

$$\lim_{s\to\infty} \frac{\mathbb{P}(S > s)}{\mathbb{P}(X_1 > s)} = \rho.$$

This next result makes the relationship between regular variation and extreme Value-at-Risk explicit

Lemma 5.3 (cf Lemma 2.3 of Embrechts et al., 2009)**.** *If* $\boldsymbol{X} \in \mathrm{MRV}_+^d(-1/\xi)$*. Then*

$$\lim_{\alpha\nearrow 1} \frac{\mathrm{VaR}_\alpha(S)}{\mathrm{VaR}_\alpha(X_1)} = \rho^\xi.$$

Hence, the constant ρ determines the extreme Value-at-Risk (expressed as a limit) for the sum of losses $S = X_1 + X_2 + \cdots + X_d$, normalized by the common marginal. It was shown in Barbe et al. (2006) (see also Theorem 4.1 of Embrechts et al., 2009) that $\boldsymbol{X} \in \mathrm{MRV}_+^d(-1/\xi)$ implies

$$\rho \equiv \rho(H,\xi) := \int_{\mathbb{S}_+^{d-1}} (u_1^\xi + u_2^\xi + \cdots + u_d^\xi)^{1/\xi} H(d\boldsymbol{u}) \tag{5.2}$$

where H is the spectral measure in (5.1) which, theoretically, could be any finite measure on $\mathbb{S}_+^{d-1}$ satisfying marginal equality constraints

$$1 = \int_{\mathbb{S}_+^{d-1}} u_j H(d\boldsymbol{u}),\ j = 1,\ldots,d. \tag{5.3}$$

Well known universal bounds on the value of ρ are given by

$$d \leq \rho(H,\xi) \leq d^{1/\xi} \quad \xi \leq 1 \tag{5.4}$$

$$d^{1/\xi} \leq \rho(H,\xi) \leq d \qquad \xi \geq 1, \tag{5.5}$$

(see e.g. Corollary 4.2 of Embrechts et al., 2009). Observe that for $\xi = 1$ we have $\rho = d$, regardless of the form of H. Otherwise, $\rho = d$ corresponds to mutual independence

and $\rho = d^{1/\xi}$ corresponds to complete tail dependence of components of the vector $\boldsymbol{X}$. The possibility of ρ taking any value within the closed interval $[d \wedge d^{1/\xi}, d \vee d^{1/\xi}]$ reflects our lack of knowledge about the tail dependence structure of $\boldsymbol{X}$, or equivalently, its spectral measure H. The fact that H is itself an infinite dimensional parameter, which may vary within a large class of measures makes fully characterizing tail dependence a difficult problem (see Einmahl et al., 2012, and references therein). In contrast, one can estimate, in practice, various finite dimensional functionals which summarize the dependence of $\boldsymbol{X}$. One popular such set of functionals is the *extremal coefficients* (see e.g Smith, 1990, Schlather and Tawn, 2002 and Cooley et al., 2006).

Definition 5.4 (Extremal coefficient)**.** Let H be the spectral measure of $\boldsymbol{X} \in \mathrm{MRV}_+^d(-1/\xi)$ and J a non-empty subset of $\{1, \ldots, d\}$. The J^{th} extremal coefficient of with respect to H is

$$\vartheta_H(J) := \int_{\mathbb{S}_+^{d-1}} \max_{j \in J}\{u_j\} H(d\boldsymbol{u}).$$

A collection of non-negative constants $\boldsymbol{c} = (c_J)_{J \subset \{1,\ldots,d\}} \in \mathbb{R}_+^{2^d-1}$ is called a *consistent* set of extremal coefficients if

$$\boldsymbol{c} \in \Theta := \left\{ \boldsymbol{\vartheta} \in \mathbb{R}_+^{2^d-1} : \sum_{L \subset J} (-1)^{|L|+1} \vartheta_J \geq 0, \text{ for all } J \subset \{1, \ldots, d\} \right\}.$$

Extremal coefficients alone do not fully characterize the spectral measure, except in special cases (Strokorb and Schlather, 2013). However, to the best of our knowledge it has not been determined prior to this work, the extent to which additional information given by extremal coefficients constrain the range of possible ρ. This is precisely the motivation for this research. Our objective is to determine sharp bounds on the value of ρ when obtaining full or partial knowledge of the extremal coefficients. That is, we

want to determine exactly the interval

$$(P^*) \qquad \left(\inf_H \rho(H,\xi), \sup_H \rho(H,\xi)\right) \tag{5.6}$$

$$\text{subject to:} \qquad \int_{\mathbb{S}_+^{d-1}} \max_{j\in J}\{u_j\} H(d\boldsymbol{u}) = c_J, \text{ for all } J \in \mathcal{J}, \tag{5.7}$$

where the supremum and infimum are taken over all finite measures on $\mathbb{S}_+^{d-1}$, and $\mathcal{J}$ is a given collection of subsets of $\{1, 2, \ldots, d\}$.

Assumption 5.5. *We assume that the marginal constraints* (5.3) *are always included in* (5.7) *by insuring that the singletons* $\{1\}, \ldots, \{d\}$ *belong to* $\mathcal{J}$ *and* $c_{\{j\}} = 1$ *for* $j = 1, \ldots, d$. *To avoid further trivialities also assume* $\mathcal{J}$ *is sufficiently rich such that*

$$1 = \sum_{j=1}^{d} u_j < \sum_{J\in\mathcal{J}} \max_{j\in J}\{u_j\}, \text{ for all } \boldsymbol{u} \in \mathbb{S}_+^{d-1}.$$

In particular, this holds if $\mathcal{J}$ *includes all pairs or* $\{1, \ldots, d\} \in \mathcal{J}$.

Our contribution is twofold. First, we characterize the solution to the problem (P^*). We show that the inf and sup in (5.6) are in fact attained by discrete measures that are supported on a finite set of atoms. In each case, the number of atoms is not more than the number of constraints in (5.7). Our results are established using the theory of *semi-infinite programming* (SIP) and in particular *linear* SIP (LSIP) (see e.g. Shapiro, 2009 and Goberna and Lopez, 1998). These theoretical results, while enlightening, do not readily provide practical algorithms to obtain the bounds in (5.6), which is of utmost interest in applications.

Our second contribution focuses on an algorithm for (P^*). We prove that the upper and lower bounds (5.6) are attained by solving a pair of convex optimization problems in finite dimension. Hence, ϵ-precision estimates of the bounds on ρ are possible to obtain in a finite number of steps. We demonstrate theoretically and empirically that in some cases, very limited information about the extremal coefficients (e.g.

only pair-wise extremal coefficients) can lead to significant reduction in the range of possible values for ρ. This shows that significant information about the possible range of extreme VaR could be derived from readily estimable parameters (extremal coefficients). The scaling of the optimization problems to large dimensional portfolios ($d \approx 50$) remains a challenge, nevertheless our structural results and experiments in $d \leq 8$ suggest that promising results are possible with high performance computing or more sophisticated optimization techniques.

5.2 Linear semi-infinite programming

The purpose of this section is to establish definitions and notations from linear semi-infinite programming that we will use throughout this Chapter. We assume the reader has some background in mathematical programming, a working knowledge of linear programming and Lagrangian duality. Some additional results are presented in Section 5.7.

Linear semi-infinite programs are mathematical programs that can be formulated in the following way

$$\begin{aligned} (P) \qquad & \inf_{\boldsymbol{x}\in\mathbb{R}^p} \quad \boldsymbol{c}^\top \boldsymbol{x} \\ & \text{subject to:} \quad b(t) - \boldsymbol{a}(t)^\top \boldsymbol{x} \leq 0,\ t \in T, \end{aligned}$$

for functions $\boldsymbol{a} : T \mapsto \mathbb{R}^p$, $b : T \mapsto \mathbb{R}$, and T is a (possibly infinite) index set. For a given mathematical program, say $(\tilde{P})$, we use the notation $\text{val}(\tilde{P})$, to denote its optimal value while $\text{sol}(\tilde{P})$ denotes the solution set, i.e. the set of feasible points that yield optimal values. Generally, $\text{val}(\tilde{P})$ may be infinite and $\text{sol}(\tilde{P})$ my be empty. If $\text{sol}(\tilde{P}) = \varnothing$, then by convention $\text{val}(\tilde{P}) = \infty$ and we say $(\tilde{P})$ is *unsolvable*. It is said

that the *Slater condition* holds for (P) if there exists $\tilde{\boldsymbol{x}} \in \mathbb{R}^p$ such that

$$b(t) - \boldsymbol{a}(t)^\top \tilde{\boldsymbol{x}} < 0, \text{ for all } t \in T.$$

We make the following assumption throughout

Assumption 5.6. *T is a compact subset of $\mathbb{R}^d$ and $\boldsymbol{a} : T \mapsto \mathbb{R}^p$, $b : T \mapsto \mathbb{R}$ are bounded and continuous on T.*

Define the *Lagrangian* of problem (P) as the function $L : \mathbb{R}^p \times \Omega \mapsto \mathbb{R}$

$$L(\boldsymbol{x}, \omega) = \boldsymbol{c}^\top \boldsymbol{x} + \int_T \left(b(t) - \boldsymbol{a}(t)^\top \boldsymbol{x}\right) \omega(dt), \tag{5.8}$$

where Ω is the space of finite (non-negative) Borel measures on T.

Remark 5.7. It is not always common in LSIP to make the Assumption 5.6 but doing so here allows us to express the Lagrangian function as (5.8). This follows from the fact that the topological *dual space* of continuous functions on the compact set $T \subset \mathbb{R}^p$ is indeed the space of Borel measures on T. This shows that the problem (P^*) on page 108 falls within the framework of linear semi-infinite programming. The interested reader is referred to Ch. 2 of Goberna and Lopez (1998) for additional background.

We define the *dual function* $g : \Omega \mapsto \mathbb{R}$ as

$$g(\omega) = \inf_{\boldsymbol{x} \in \mathbb{R}^p} L(\boldsymbol{x}, \omega).$$

The dual function yields a lower bound on the optimal value of (P) since for any feasible $\tilde{\boldsymbol{x}} \in \mathbb{R}^p$, it follows that

$$\int_T \left(b(t) - \boldsymbol{a}(t)^\top \tilde{\boldsymbol{x}}\right) \omega(dt) \leq 0,$$

which implies

$$g(\omega) = \inf_{\boldsymbol{x}\in\mathbb{R}^p} L(\boldsymbol{x},\omega) \leq \boldsymbol{c}^\top\tilde{\boldsymbol{x}} + \int_T \left(b(t) - \boldsymbol{a}(t)^\top\tilde{\boldsymbol{x}}\right)\omega(dt) \leq \boldsymbol{c}^\top\tilde{\boldsymbol{x}}. \tag{5.9}$$

The fact that $\tilde{\boldsymbol{x}}$ was arbitrarily feasible implies $g(\omega) \leq \text{val}(P)$. This inequality is superficial unless $\int_T \boldsymbol{a}(t)\omega(dt) = \boldsymbol{c}$, otherwise Assumption 5.6 implies that $g(\omega) = \inf_{\boldsymbol{x}\in\mathbb{R}^p} L(\boldsymbol{x},\omega)$ is unbounded below in $\boldsymbol{x}$. Any measure $\omega \in \Omega$ for which $\int_T \boldsymbol{a}(t)\omega(dt) = \boldsymbol{c}$ holds is referred to as *dual feasible*. Thus we arrive at the following *dual problem*

$$\begin{aligned} (D) \qquad \sup_{\omega\in\Omega} \quad & \int_T b(t)\omega(dt) \\ \text{subject to:} \quad & \int_T \boldsymbol{a}(t)\omega(dt) = \boldsymbol{c}. \end{aligned}$$

A common task with many optimization problems is to determine the existence (or non-existence) of a *duality gap,* $|\text{val}(P) - \text{val}(D)|$. If $\text{val}(P) = \text{val}(D)$, then it suffices to solve either (P) or (D) to obtain the optimal value, so long as both problems are *solvable*. The condition $\text{val}(P) = \text{val}(D)$ with $\text{sol}(D) \neq \varnothing$ is known as *strong duality.*

Another important concept from LSIP theory is the possible *reducibility* of the constraint set. Consider a finite index set $T_m \subset T$ with $|T_m| \leq m$. Solving problem (P) when the constraints are restricted to the finite set T_m reduces to a standard linear program

$$\begin{aligned} (P_m) \qquad \inf_{\boldsymbol{x}\in\mathbb{R}^p} \quad & \boldsymbol{c}^\top\boldsymbol{x} \\ \text{subject to:} \quad & b(t_i) - \boldsymbol{a}(t_i)^\top\boldsymbol{x} \leq 0,\ t_i \in T_m,\ i = 1,\ldots,m, \end{aligned}$$

which yields the corresponding dual

$$(D_m) \qquad \sup_{\boldsymbol{\omega}\in\mathbb{R}^m_+} \quad \sum_{i=1}^{m} b(t_i)\omega_i$$
$$\text{subject to:} \quad \sum_{i=1}^{m} \boldsymbol{a}(t_i)\omega_i = \boldsymbol{c},\ t_i \in T_m,\ i = 1, \ldots, m.$$

Problem (P_m) is called a *discretization* of (P). The feasible set for (P) is contained in the feasible set for (P_m). Hence, $\text{val}(P) \leq \text{val}(P_m)$. If for any $\varepsilon > 0$, there exists (P_m) such that $\text{val}(P_m) - \text{val}(P) \leq \varepsilon$ than we say (P) is *discretizable.* If there exists (P_m) such that $\text{val}(P_m) = \text{val}(P)$ then (P) is said to be *reducible.*

5.3 Main results

We are now ready to present our main results. From (5.4) and (5.5) it is seen that the tail index ξ plays a pivotal role in determining the range of possible ρ. Consequently, in order to handle all variants of (P^*), we need to consider a total of four optimization problems depending on the value of ξ and wether we are minimizing or maximizing. To this end, consider problems (I), (II), (III) and (IV) as classified by the following scheme

Optimization Scheme

$$(P^*) \qquad \sup_H \left(\inf_H\right) \int_{\mathbb{S}_+^{d-1}} \left(u_1^\xi + u_2^\xi + \cdots + u_d^\xi\right)^{1/\xi} H(d\boldsymbol{u})$$

$$\text{subject to: } \left\{ \int_{\mathbb{S}_+^{d-1}} \max_{j\in J}\{u_j\} H(d\boldsymbol{u}) = c_J \right\}_{J\in\mathcal{J}},$$

	$\xi \le 1$	$\xi \ge 1$
$\sup_H$	IV Conic Program	I Linear Program
$(\inf_H)$	III Linear Program	II Conic Program

We will show that the optimal value for all four problems (I) $\cdots$ (IV) can be obtained by solving convex optimization problems of *finite* dimension. The solutions to Problems (I) and (III) involve solving a linear program, while (II) and (IV) involve conic programming (see Section 5.6). To arrive at these results, we first prove that (I) $\cdots$ (IV) are indeed LSIPs. We continue the notation from Section 5.2 by denoting $\text{val}(\cdot), \text{sol}(\cdot)$ as the optimal value and solution set for each problem (I) $\ldots$ (IV).

Theorem 5.8. *Suppose Assumption 5.5 holds. There exists linear semi-infinite programs* $(\text{I}'), (\text{II}'), (\text{III}')$ *and* (IV') *such that*

(i) Assumption 5.6 is satisfied.

(ii) The Slater condition holds.

(iii) The optimal value is finite.

(iv) Strong duality holds for each pair $(\text{I}, \text{I}'), (\text{II}, \text{II}'), (\text{III}, \text{III}')$ *and* (IV, IV').

Proof. We content ourselves with proving the case (I) as proofs for (II), ..., (IV) are very similar. Fix $\xi \geq 1$. Let $p = |\mathcal{J}|$ and $\boldsymbol{c} = (c_J)_{J\in\mathcal{J}} \in \mathbb{R}^p_+$. Define the continuous functions $b : \mathbb{S}^{d-1}_+ \mapsto \mathbb{R}_+$, $\boldsymbol{a} : \mathbb{S}^{d-1}_+ \mapsto \mathbb{R}^p_+$

$$b(\boldsymbol{u}) = \left(u_1^\xi + \cdots + u_d^\xi\right)^{1/\xi}$$

$$\boldsymbol{a}(\boldsymbol{u}) = \left(\max_{j\in J}\{u_j\}\right)_{J\in\mathcal{J}}.$$

The mapping $(\boldsymbol{u}, \boldsymbol{x}) \mapsto b(\boldsymbol{u}) - \boldsymbol{a}(\boldsymbol{u})^\top \boldsymbol{x}$ is linear in $\boldsymbol{x}$ and hence, the following problem (I′) is a linear semi-infinite program

$$\begin{array}{lll} (\mathrm{I}') & \inf\limits_{\boldsymbol{x}\in\mathbb{R}^p} & \boldsymbol{c}^\top \boldsymbol{x} \\ & \text{subject to:} & b(\boldsymbol{u}) - \boldsymbol{a}(\boldsymbol{u})^\top \boldsymbol{x} \leq 0,\ \boldsymbol{u} \in \mathbb{S}^{d-1}_+. \end{array}$$

Since $\mathbb{S}^{d-1}_+ \subset \mathbb{R}^d$ is compact and the functions b and $\boldsymbol{a}$ are continuous and bounded on $\mathbb{S}^{d-1}_+$, Assumption 5.6 is satisfied for (I′). Letting $\mathcal{H}$ denote the space of finite Borel measures on $\mathbb{S}^{d-1}_+$, by the Lagrangian duality illustrated in Section 5.2, it follows that the dual of (I′) is

$$\begin{array}{lll} (\mathrm{I}) & \sup\limits_{H\in\mathcal{H}} & \displaystyle\int_{\mathbb{S}^{d-1}_+} \left(u_1^\xi + \cdots + u_d^\xi\right)^{1/\xi} H(d\boldsymbol{u}) \\ & \text{subject to:} & \left\{\displaystyle\int_{\mathbb{S}^{d-1}_+} \max_{j\in J}\{u_j\} H(d\boldsymbol{u}) = c_J\right\}_{J\in\mathcal{J}}. \end{array}$$

Observe that $\tilde{\boldsymbol{x}} \equiv \mathbf{1} \in \mathbb{R}^p$ is primal feasible since

$$b(\boldsymbol{u}) = \left(u_1^\xi + \cdots + u_d^\xi\right)^{1/\xi} \leq \sum_{j=1}^d u_j < \sum_{J\in\mathcal{J}} \max_{j\in J}\{u_j\} = \boldsymbol{a}(\boldsymbol{u})^\top \mathbf{1},\ \text{for all } \boldsymbol{u} \in \mathbb{S}^{d-1}_+, \tag{5.10}$$

where the second inequality in (5.10) holds by Assumption 5.5. Hence the Slater

condition holds. This, in view of (5.5) and (5.9) implies that

$$-\infty < d^{1/\xi} \leq \mathrm{val}(\mathrm{I}') \leq \sum_{J\in\mathcal{J}} c_J < \infty.$$

Thus it follows from Proposition 5.15 of Section 5.7 that sol(I) is non-empty and $\mathrm{val}(\mathrm{I}') = \mathrm{val}(\mathrm{I})$. This completes the proof. □

The consequences of Theorem 5.8 is not only the establishment of a linear semi-infinite program for each problem in (P^*) but moreover each one of (II), ..., (IV) satisfies sufficient conditions for reducibility. Indeed we have the following non-trivial result which shows that optimal measures exist for (P^*) that are concentrated on at most $|\mathcal{J}|$ atoms.

Theorem 5.9. *Let* $A = (a_{jK})_{j=1,K\in\mathcal{J}}^{j=d} \in \mathbb{R}_+^d \times \mathbb{R}_+^{|\mathcal{J}|}$ *be a non-negative matrix and define*

$$\mathcal{A}_{\boldsymbol{c}} := \left\{ A \in \mathbb{R}_+^d \times \mathbb{R}_+^{|\mathcal{J}|} : \sum_{K\in\mathcal{J}} \max_{j\in J}\{a_{jK}\} = c_J, J \in \mathcal{J} \right\}. \tag{5.11}$$

Then

$$\sup_{A\in\mathcal{A}_{\boldsymbol{c}}} \sum_{K\in\mathcal{J}} \left(a_{1K}^{\xi} + a_{2K}^{\xi} + \cdots + a_{dK}^{\xi} \right)^{1/\xi} = \begin{cases} \mathrm{val}(\mathrm{I}) & \xi \geq 1 \\ \mathrm{val}(\mathrm{IV}) & \xi \leq 1, \end{cases}$$

and

$$\inf_{A\in\mathcal{A}_{\boldsymbol{c}}} \sum_{K\in\mathcal{J}} \left(a_{1K}^{\xi} + a_{2K}^{\xi} + \cdots + a_{dK}^{\xi} \right)^{1/\xi} = \begin{cases} \mathrm{val}(\mathrm{III}) & \xi \leq 1 \\ \mathrm{val}(\mathrm{II}) & \xi \geq 1. \end{cases}$$

Proof. By Theorem 5.8 we have for (I′) that Assumption 5.6 holds, the Slater condition holds and val(I′) is finite. Hence, by Proposition 5.17, there must exist a discretization (I_m) with $m \leq |\mathcal{J}|$ such that $\mathrm{val}(\mathrm{I}_m) = \mathrm{val}(\mathrm{I}) < \infty$. The last statement

means that

$$\begin{aligned}\text{val(I)} = \quad & \sup_{\substack{\boldsymbol{u}_k \in \mathbb{S}_+^{d-1} \\ h_k \geq 0}} \sum_{k=1}^{m} \left(u_{1k}^{\xi} + \cdots + u_{dk}^{\xi} \right)^{1/\xi} h_k \\ & \text{subject to : } \left\{ \sum_{k=1}^{m} \max_{j \in J} \{u_{jk}\} h_k = c_J \right\}_{J \in \mathcal{J}}.\end{aligned}$$

making the change of variables $a_{jk} = u_{jk} h_k$ gives

$$\begin{aligned}\text{val(I)} = \quad & \sup_{a_{jk} \geq 0} \sum_{k=1}^{m} \left(a_{1k}^{\xi} + \cdots + a_{dk}^{\xi} \right)^{1/\xi} \\ & \text{subject to : } \left\{ \sum_{k=1}^{m} \max_{j \in J} \{a_{jk}\} = c_J \right\}_{J \in \mathcal{J}}.\end{aligned}$$

Thus we have proved the result for (I). The cases (II), . . . , (IV) are redundant. □

It follows from Theorem 5.9 that val(II) and val(IV) can be obtained by solving convex optimization problems.

Corollary 5.10. *Theorem 5.9 implies*

$$\begin{aligned}\text{val(II)} \quad = \quad & \inf_{\substack{A \in \mathcal{A}_{\boldsymbol{c}} \\ \boldsymbol{t} \in \mathbb{R}^{|\mathcal{J}|}}} \sum_{J \in \mathcal{J}} t_J, \ \xi \geq 1, \qquad (5.12) \\ & \text{subject to : } \left\{ \left(a_{1J}^{\xi} + a_{2J}^{\xi} + \cdots + a_{dJ}^{\xi} \right)^{1/\xi} \leq t_J \right\}_{J \in \mathcal{J}}. \\ \text{val(IV)} \quad = \quad & \sup_{\substack{A \in \mathcal{A}_{\boldsymbol{c}} \\ \boldsymbol{t} \in \mathbb{R}^{|\mathcal{J}|}}} \sum_{J \in \mathcal{J}} t_J, \ \xi \leq 1, \qquad (5.13) \\ & \text{subject to : } \left\{ \left(a_{1J}^{\xi} + a_{2J}^{\xi} + \cdots + a_{dJ}^{\xi} \right)^{1/\xi} \geq t_J \right\}_{J \in \mathcal{J}}.\end{aligned}$$

Problems (5.12) and (5.13) are *conic programs* (see Section 5.6). A survey of existing software for conic programs did not yield any solvers that allow input of (5.12) and (5.13) directly. In particular there does not appear to be implementations that natively handle the max-linear constraints (5.11). However, if a-priori we fix the

permutations $\{\pi_J\}_{J\in\mathcal{J}}$ such that

$$a_{\pi_J(1)J} \leq a_{\pi_J(2)J} \leq \cdots \leq a_{\pi_J(d)J},\ J \in \mathcal{J},$$

then there always exists a (sparse) pattern matrix $Q \in \{0,1\}^{|\mathcal{J}|} \times \{0,1\}^{d|\mathcal{J}|}$ such that the condition $Q\boldsymbol{a} = \boldsymbol{c}$ implies $A \in \mathcal{A}_{\boldsymbol{c}}$, where $\boldsymbol{a}$ is the vectorized entries of A. As far as we know, finding a sufficient pattern matrix Q, given a partial set of extremal coefficients $\{\vartheta_J, J \in \mathcal{J}\}$ is an interesting an open problem. Alternatively, one can always saturate the optimization problems (5.12) and (5.13) with all $d!$ permutations, taking the problem from $(d+1)|\mathcal{J}|$ parameters to $(d+1)!$. This strategy is employed in Section 5.4 to make use of off-the-shelf solvers and is feasible for moderate dimension $d \leq 10$. An alternative approach is to develop algorithms that handle max-linear constraints natively, which is a subject of current research.

Ignoring the technicalities of specifying max-linear constraints, when $\xi = 2$, relation (5.12) corresponds to a *second order cone program* (SOCP) and efficient solvers for SOCPs are readily available. More generally, when $\xi > 1$, the problem (5.12) was studied in detail by Xue and Ye (2000). The authors therein developed an efficient interior-point algorithm that solves (5.12) in no more than a factor $\sqrt{d}$ time over standard SOCPs. For the case (5.13), where $\xi < 1$, we could not find an existing efficient algorithm that readily applies. However, due to the similar geometric structure of (5.12) and (5.13) we speculate that strategies following that of Xue and Ye (2000) can be explored to yield an efficient solver. In any case one may always apply general non-linear methods for convex programming that will provide an ϵ-precision solution in polynomial time.

Turning our attention to (I) and (III), the optimizations for val(I) and val(III) appearing in Theorem 5.9 are non-convex. This appears at first to be a set-back, however, in this case, obtaining the optimal values val(I) and val(III) corresponds

to solving a standard (finite dimensional) linear program. Intuition comes from the fact that when one wants to minimize a concave function over a convex set, then the solutions should appear at certain *extreme points* of the (convex) feasible region. The following theorem shows that indeed this is the case for problems (I) and (III).

Theorem 5.11. *The optimal values of* (I) *and* (III) *are obtained by solving the linear programs*

$$\begin{aligned} \text{val(I)} &= \sup_{\boldsymbol{\beta}\in\mathbb{R}_+^{2^d-1}} \sum_{J\subset\{1,\ldots,d\}\setminus\varnothing} |J|^{1/\xi}\beta_J,\ \xi \geq 1, && (5.14)\\ \text{val(III)} &= \inf_{\boldsymbol{\beta}\in\mathbb{R}_+^{2^d-1}} \sum_{J\subset\{1,\ldots,d\}\setminus\varnothing} |J|^{1/\xi}\beta_J,\ \xi \leq 1, && (5.15)\\ &\quad \text{subject to}: \left\{ \sum_{K\subset\{1,\ldots,d\}\setminus\varnothing} \mathbb{I}\left\{(K\cap J)\neq\varnothing\right\}\beta_K = c_J \right\}_{J\in\mathcal{J}}. \end{aligned}$$

A proof of this result requires several steps and two lemma. For this reason we postpone the proof for Section 5.8.

Remark 5.12. Relations (5.14) and (5.15) imply that optimal spectral measures associated with the optimal values are finitely supported on the set

$$\left\{ |J|^{-1}\left(\mathbf{1}_J(j)\right)_{j=1}^d : J\subset\{1,\ldots,d\} \right\} \subset \mathbb{S}_+^{d-1}$$

with mass $|J|\beta_J$ at each atom. Such a spectral measure corresponds to the *Tawn-Molchanov* max-stable model (Strokorb and Schlather, 2013), which is the spectral measure uniquely characterized by a consistent set of extremal coefficients.

To summarize, we have shown that indeed val(I) $\cdots$ val(IV) can be obtained by solving convex optimization problems of finite dimension and in-particular, efficient-interior point methods already exist for val(I) $\cdots$ val(III), while an efficient algorithm for val(IV) is conjectured.

5.4 Examples

In this section, we demonstrate using two examples, the impact of the theory given in Section 5.3. In the first example we illustrate with numerical experiments, the amount of information for extreme VaR given by all bivariate extremal coefficients. The second example fully characterizes the case when given only the extremal coefficient associated with the complete set $D = \{1, \dots, d\}$.

5.4.1 Example: bi-variate constraints

Here we consider the case where all bi-variate extremal coefficients are given. Our general procedure is to solve the optimization problems (I) $\dots$, (IV) where the pairwise extremal coefficients are randomly generated. We define the information about extreme VaR as the reduction in the width of the interval given by universal bounds for ρ^ξ

$$\mathcal{I}(\boldsymbol{\vartheta}) := \begin{cases} 1 - \frac{|\text{val(I)} - \text{val(II)}|}{|d - d^\xi|} & \xi \geq 1 \\ 1 - \frac{|\text{val(III)} - \text{val(IV)}|}{|d - d^\xi|} & \xi \leq 1. \end{cases} \tag{5.16}$$

In order to generate a variety of consistent pairwise extremal coefficients, we first generate random spectral measures using a sampling procedure that enables a large range of extremal coefficients. For simplicity, we generate finitely supported discrete spectral measures since they are dense in the class of all valid spectral measures and their extremal coefficients can be readily computed. Discrete spectral measures with finite support are fully characterized by non-negative matrices with rows that sum to unity. If the matrices themselves are dense, then the resulting spectral measure concentrates mass within the interior of $\mathbb{S}_+^{d-1}$. Therefore, to encourage a variety of extremal coefficients we impose sparsity by setting roughly 70% of entries in the generated matrix to zero. The sampling procedure is described as follows:

Spectral Measure Generation

1. Draw number of atoms $m \sim \text{Poisson}(d)$.

2. For each $k = 1, \ldots, m$, $j = 1, \ldots, d$, draw $B \sim \text{Bernoulli}(0.3)$, $U \sim \text{uniform}(0, 1)$ and set $\tilde{a}_{jk} = UB$.

3. Set $A = (a_{jk})_{d \times m}$, where $a_{jk} = \tilde{a}_{jk} / \sum_{k=1}^{m} \tilde{a}_{jk}$.

Each matrix A generated by the above procedure results in a valid spectral measure. Moreover, the procedure yields a wide variety of sets of consistent extremal coefficients

$$\vartheta_J = \sum_{k=1}^{m} \max_{j \in J} \{a_{jk}\}, \ \varnothing \neq J \subset \{1, \ldots, d\}.$$

The following procedure describes our numerical experiment

Simulation Procedure

Input d, ξ

For each iteration $i = 1, \ldots, N$

1. Generate $A^{(i)} = (a_{jk}^{(i)})_{d \times m}$ using the spectral measure generation procedure above.

2. Calculate the pairwise extremal coefficients

$$\vartheta_J^{(i)} = \sum_{k=1}^{m} \max_{j \in J} \{a_{jk}^{(i)}\}, \ \varnothing \neq J \subset D, |J| = 2.$$

3. Solve the optimization problems $(\mathrm{I}, \mathrm{II})$ or $(\mathrm{III}, \mathrm{IV})$ via linear programming and conic programming.

4. Calculate the information criterion $\mathcal{I}(\boldsymbol{\vartheta}^{(i)})$ defined in relation (5.16).

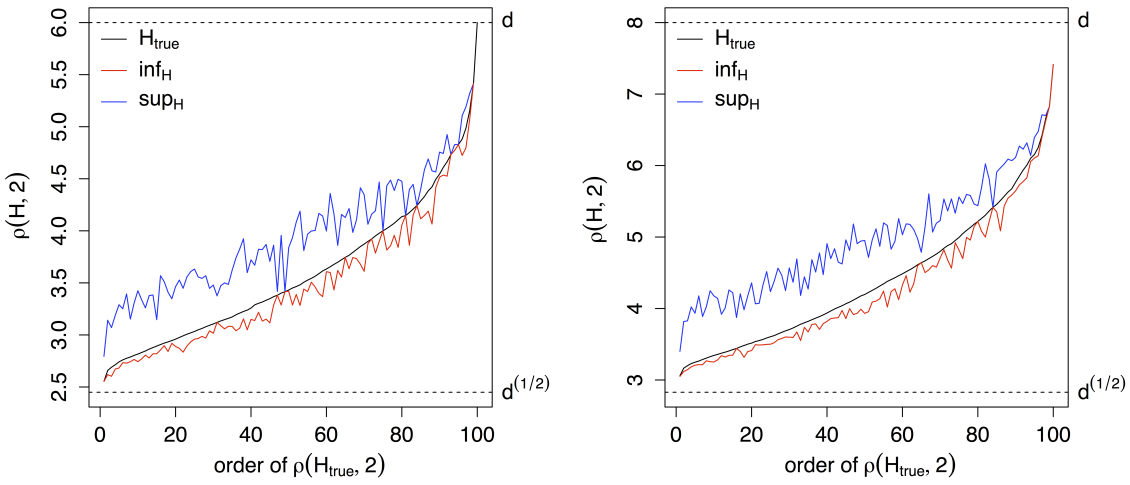
While the linear programs (5.14) and (5.15) can be solved efficiently for moderately high dimension, there is no readily available software that can handle direct input of the conic programs (5.12) and (5.13). This is due to the equality constraints being max-linear and the conic constraints being anti-symmetric, unless $\xi = 2$ (see Section 5.6). For these reasons, until a custom algorithm is implemented, we are restricted to conducting numerical experiments for lower dimensions $d \leq 8$ and where $\xi = 2$, i.e. a second-order cone program where the max-linear constraints can be handled by saturating the problem with $d!$ atoms having a-priori known ordering and therefore the constraints can be made linear through a sparse pattern matrix as described in the Section 5.3. The results of the simulation procedure are given in Table 5.1 for the cases $d = 3, 6, 8$. We report the minimum and median information $\mathcal{I}(\boldsymbol{\vartheta})$ computed

Table 5.1: VaR information $\mathcal{I}(\boldsymbol{\vartheta})$ for all pairwise extremal coefficients.

d	ξ	$\lvert d^{\xi} - d\rvert$	$\min_{i=1,\dots,n} \mathcal{I}(\boldsymbol{\vartheta}^{(i)})$	$\text{median}_{i=1,\dots,n} \mathcal{I}(\boldsymbol{\vartheta}^{(i)})$
3	2.0	6	0.78	0.88
6	2.0	20	0.78	0.88
8	2.0	56	0.79	0.89

over $n = 100$ replications. In Figure 5.1 we visually display the width of the bounds with respect to the dependence bounds $(d^{1/\xi}, d)$. To solve the SOCP we deployed the commercial solver Gurobi (Gurobi Optimization, 2015). To give a sense of the computation required, optimizations for a single replication in dimension $d = 8$, which contains $(d+1)! = 362{,}880$ parameters, were completed on average in 25.11 iterations and had a mean completion time of 81.06 seconds on a 2-core 1.7GHz Macbook Air with 4GB RAM.

Figure 5.1: Results of two experiments. **left:** $d = 6$. **right:** $d = 8$. 100 random discrete spectral measures (H_{true}) were drawn and ordered along the x-axis according to their value of $\rho(H_{\text{true}}, \xi)$. Bounds $(\inf_H, \sup_H)$ correspond to optimization over all spectral measures with identical margins and fixed bivariate extremal coefficients.



5.4.2 Example: single d-variate constraint

Now we fully characterize the range of possible extreme VaR when given the single extremal coefficient associated with the entire set $D = \{1, \ldots, d\}$. The following results show that in this special case, exploiting the complete symmetry of the constraints yield closed form solutions of the optimal values $\text{val(I)}, \ldots, \text{val(IV)}$.

Theorem 5.13. *Consider problems* (II) *and* (IV) *when given only the d-variate extremal coefficeint*

$$\begin{aligned} \mathcal{J} &= \{\{1\}, \{2\}, \ldots, \{d\}, \{1, \ldots, d\}\} \\ \boldsymbol{c} &= (1, 1, \ldots, 1, \vartheta) \in \mathbb{R}_+^{d+1}. \end{aligned}$$

In this case, the optimal value is given by

$$\begin{aligned} v(\vartheta) &:= \left\{\vartheta^\xi + (d-1)^{1-\xi}(d-\vartheta)^\xi\right\}^{1/\xi} \\ &= \begin{cases} \inf_{\boldsymbol{u} \in \mathbb{S}_+^{d-1}} \left\{ d \left(\sum_{j=1}^d u_j^\xi \right)^{1/\xi} : \max_{j \in D}\{u_j\} = \frac{\vartheta}{d} \right\} = \text{val(II)} & \xi \geq 1 \\ \sup_{\boldsymbol{u} \in \mathbb{S}_+^{d-1}} \left\{ d \left(\sum_{j=1}^d u_j^\xi \right)^{1/\xi} : \max_{j \in D}\{u_j\} = \frac{\vartheta}{d} \right\} = \text{val(IV)} & \xi \leq 1 \end{cases}. \end{aligned} \tag{5.17}$$

Proof. The value (5.17) is obtained by a discrete measure that places unit mass on each of the vectors

$$\boldsymbol{u}_k = \frac{\vartheta}{d} \mathbb{I}_{\{k\}} + \frac{d - \vartheta}{d(d-1)} \left(\mathbf{1} - \mathbb{I}_{\{k\}} \right), j = 1, \ldots d,$$

where $\mathbf{1} = (1, 1, \ldots, 1)^\top$ and $\mathbb{I}_{\{k\}} \in \{0, 1\}^d$ is the binary vector having 1 in its jth entry and 0's elsewhere. We verify the KKT optimality conditions of Proposition 5.16

using this measure and the dual variables

$$\begin{aligned} x_k &= \frac{v(\vartheta) - \vartheta v'(\vartheta)}{d},\ k = 1, \ldots, d, \\ x_{d+1} &= v'(\vartheta). \end{aligned}$$

Dual feasibility (5.19):

$$\sum_{k=1}^{d} \max_{j \in J} \{u_{jk}\} \cdot 1 = \sum_{k=1}^{d} \max_{j \in J} \left\{ \frac{\vartheta}{d} \mathbf{1}_{J \cup \{k\}}(j), \frac{d - \vartheta}{d(d-1)} \left(1 - \mathbf{1}_{J \cup \{k\}}(j)\right) \right\} = \begin{cases} 1 & \text{if } |J| = 1 \\ \vartheta & \text{if } |J| = d \end{cases}$$

Complementary slackness (5.20):

Observe that for every $k = 1, \ldots, d$

$$\frac{1}{d} v(\vartheta) = \frac{1}{d} \left\{ \vartheta^{\xi} + (d-1)^{1-\xi} (d - \vartheta)^{\xi} \right\} = \left(u_{1k}^{\xi} + u_{2k}^{\xi} + \cdots + u_{dk}^{\xi} \right)^{1/\xi}.$$

Hence,

$$\begin{aligned} \sum_{j=1}^{d} u_{jk} x_j + \max_{j=1,\ldots,d} \{u_{jk}\} x_{d+1} &= \frac{1}{d} v(\vartheta) - \frac{\vartheta}{d} v'(\vartheta) + \frac{\vartheta}{d} v'(\vartheta) \\ &= \frac{1}{d} v(\vartheta) = \left(u_{1k}^{\xi} + u_{2k}^{\xi} + \cdots + u_{dk}^{\xi} \right)^{1/\xi}. \end{aligned}$$

Primal feasibility (5.21):

When $\xi \geq 1$ we have by definition

$$\left(u_1^{\xi} + u_2^{\xi} + \cdots + u_d^{\xi} \right)^{1/\xi} \geq \frac{1}{d} v \left(d \max_{j=1,\ldots,d} \{u_j\} \right),$$

and convexity of $v(\cdot)$ implies

$$\left(u_1^\xi + u_2^\xi + \cdots + u_d^\xi\right)^{1/\xi} \geq \frac{1}{d} v\left(d \max_{j=1,\ldots,d}\{u_j\}\right) \geq v'(\vartheta)\left[\max_{j\in D}\{u_j\} - \frac{\vartheta}{d}\right] + \frac{v(\vartheta)}{d}$$
$$= \sum_{j=1}^{d} u_j x_j + \max_{j\in D}\{u_j\} x_{d+1}.$$

This proves the result for case (II). When $\xi \leq 1$, the inequalities above are reversed which proves the result for (IV).

□

Theorem 5.14. *Let $B_k = [d(k+1)^{-1}, dk^{-1})$, $k = 1, \ldots d$. For problems* (I) *and* (III)*, the optimal value for the single d-variate constraint problem in which*

$$\begin{aligned} \mathcal{J} &= \{\{1\}, \{2\}, \ldots, \{d\}, \{1, \ldots, d\}\} \\ \boldsymbol{c} &= (1, 1, \ldots, 1, \vartheta) \in \mathbb{R}_+^{d+1}, \end{aligned}$$

is given by the piecewise linear function

$$\tau(\vartheta) := \sum_{k=1}^{d-1} \mathbf{1}_{B_k}(\vartheta)\left\{\frac{k^{1/\xi-1} - (k+1)^{1/\xi-1}}{k^{-1} - (k+1)^{-1}}\left(\vartheta - \frac{d}{k+1}\right) + d(k+1)^{1/\xi-1}\right\}$$
$$= \begin{cases} \text{val(I)} & \xi \geq 1 \\ \text{val(III)} & \xi \leq 1. \end{cases} \tag{5.18}$$

Proof. Without loss of generality, fix $k \in \{1, \ldots, d-1\}$ such that $\vartheta \in B_k$ and let

$$\lambda = \frac{\vartheta d^{-1} - (k+1)^{-1}}{k^{-1} - (k+1)^{-1}}.$$

Define a discrete measure $\tilde{H}$ that is supported on the points

$$\left\{ |J|^{-1} \left(\mathbf{1}_J(1), \ldots, \mathbf{1}_J(d) \right)^\top : J \subset D \right\} \subset \mathbb{S}_+^{d-1}$$

with mass at each point

$$\beta_J = \begin{cases} \frac{d\lambda}{k} \binom{d}{k}^{-1} & |J| = k \\ \frac{d(1-\lambda)}{k+1} \binom{d}{k+1}^{-1} & |J| = k+1 \\ 0 & |J| \notin \{k, k+1\}. \end{cases}$$

We will prove the theorem by showing that $\tilde{H}$ is a spectral measure for which the objective function attains the value $\tau(\vartheta)$ and $\tilde{H}$ meets the KKT optimality conditions for the standard linear program in Theorem 5.11. To show that the objective function achieves the optimal value, we verify

$$\begin{aligned} \int_{\mathbb{S}_+^{d-1}} \left(u_1^\xi + \cdots + u_d^\xi \right)^{1\xi} \tilde{H}(d\boldsymbol{u}) &= \sum_{J \subset D} |J|^{1/\xi} \beta_J \\ = \sum_{J \subset D} |J|^{1/\xi} \beta_J = d\lambda \sum_{\substack{J \subset D \\ |J|=k}} k^{1/\xi - 1} \binom{d}{k}^{-1} &+ d(1-\lambda) \sum_{\substack{J \subset D \\ |J|=k+1}} (k+1)^{1/\xi - 1} \binom{d}{k+1}^{-1} \\ &= d \left\{ \lambda k^{1/\xi - 1} + (1-\lambda)(k+1)^{1/\xi - 1} \right\} = \tau(\vartheta). \end{aligned}$$

Now define the following *dual variables* for each $K \in \mathcal{J}$

$$x_K = \begin{cases} (k+1)^{1/\xi} - k^{1/\xi} & |K| = 1 \\ \frac{k^{1/\xi - 1} - (k+1)^{1/\xi - 1}}{k^{-1} - (k+1)^{-1}} & |K| = d \end{cases}$$

To verify optimality (KKT) conditions for the standard linear program given in Theorem 5.11, we need to show

Primal feasibility:

$$\sum_{J\subset D\setminus\varnothing} \mathbb{I}\{K\cap J\neq\varnothing\}\beta_J = \begin{cases} 1 & |K|=1 \\ \vartheta & |K|=d. \end{cases}$$

When $|K|=d$ we have

$$\begin{aligned}\sum_{J\subset D\setminus\varnothing} \mathbb{I}\{K\cap J\neq\varnothing\}\beta_J &= d\lambda \sum_{\substack{J\subset D\\|J|=k}} k^{-1}\binom{d}{k}^{-1} + d(1-\lambda)\sum_{\substack{J\subset D\\|J|=k+1}} (k+1)^{-1}\binom{d}{k+1}^{-1} \\ &= \frac{\vartheta d^{-1}-(k+1)^{-1}}{k^{-1}-(k+1)^{-1}} dk^{-1} + \left(1-\frac{\vartheta d^{-1}-(k+1)^{-1}}{k^{-1}-(k+1)^{-1}}\right) d(k+1)^{-1} \\ &= \frac{\vartheta - d(k+1)^{-1}}{k^{-1}-(k+1)^{-1}}\left(k^{-1}-(k+1)^{-1}\right) + d(k+1)^{-1} = \vartheta.\end{aligned}$$

When $|K|=1$, let $j\in D$ be arbitrary, then

$$\begin{aligned}\sum_{\substack{J\subset D\setminus\varnothing\\|J|=1}} \mathbb{I}\{J\cap\{j\}\neq\varnothing\}\beta_J &= \sum_{\substack{J\subset D\setminus\varnothing\\|J|=1}} \mathbb{I}\{J\cap\{j\}\neq\varnothing\}\frac{d}{k}\binom{d}{k}^{-1} \\ &= \binom{d-1}{k-1}^{-1}\sum_{\substack{J\in 2^D\setminus\varnothing\\|J|=1}} \mathbb{I}\{J\cap\{j\}\neq\varnothing\} = 1.\end{aligned}$$

Complementary Slackness:

$$\left(|J|^{1/\xi} - \sum_{K\in\mathcal{J}} \mathbb{I}\{J\cap K\neq\varnothing\}x_K\right)\beta_J = 0.$$

It suffices to show $|J|^{1/\xi} - \sum_{K\in\mathcal{J}} \mathbb{I}\{J\cap K\neq\varnothing\}x_K = 0$ for $|J|\in\{k,k+1\}$.

When $|J| = k$ we have

$$\begin{aligned} k^{1/\xi} - \sum_{K \in \mathcal{J}} \mathbb{I}\{J \cap K \neq \varnothing\} x_K \\ &= k^{1/\xi} - k\left[(k+1)^{1/\xi} - k^{1/\xi}\right] - \frac{k^{1/\xi - 1} - (k+1)^{1/\xi - 1}}{k^{-1} - (k+1)^{-1}} \\ &= \frac{k^{1/\xi - 1} - \frac{k^{1/\xi}}{k+1} - \left(1 - \frac{k}{k+1}\right)\left[(k+1)^{1/\xi} - k^{1/\xi}\right] - k^{1/\xi - 1} + (k+1)^{1/\xi - 1}}{k^{-1} - (k+1)^{-1}}. \end{aligned}$$

Multiplying the numerator above by $k+1$ gives

$$(k+1)k^{1/\xi - 1} - k^{1/\xi} - (k+1)^{1/\xi} + k^{1/\xi} - (k+1)k^{1/\xi - 1} + (k+1)^{1/\xi} = 0.$$

The same can be shown for $|J| = k+1$.

Dual feasibility:

$$\begin{aligned} |J|^{1/\xi} - \sum_{K \in \mathcal{J}} \mathbb{I}\{J \cap K \neq \varnothing\} x_K &\leq 0 \quad \text{if } \xi \geq 1 \\ |J|^{1/\xi} - \sum_{K \in \mathcal{J}} \mathbb{I}\{J \cap K \neq \varnothing\} x_K &\geq 0 \quad \text{if } \xi \leq 1 \end{aligned}$$

When $\xi \geq 1$, the function $m \mapsto m^{1/\xi}$ is concave and hence for any $m, m_1 \leq m_2 \in \mathbb{R}_+$ such that $m \notin (m_1, m_2)$ it follows that

$$m^{1/\xi} \leq \frac{m_2^{1/\xi} - m_1^{1/\xi}}{m_2 - m_1}(m - m_1) + m_1^{1/\xi}.$$

This implies that for any $|J| \notin (k, k+1)$

$$
\begin{aligned}
|J|^{1/\xi} &\leq \frac{(k+1)^{1/\xi} - k^{1/\xi}}{k+1-k}(|J| - k) + k^{1/\xi} \\
&= |J| \left[(k+1)^{1/\xi} - k^{1/\xi}\right] + k^{1/\xi} - k\left[(k+1)^{1/\xi} - k^{1/\xi}\right] \\
&= |J| \left[(k+1)^{1/\xi} - k^{1/\xi}\right] + \frac{k^{1/\xi - 1} - (k+1)^{1/\xi - 1}}{k^{-1} - (k+1)^{-1}} = \sum_{K \in \mathcal{J}} \mathbb{I}\{J \cap K \neq \varnothing\} x_K,
\end{aligned}
$$

where the second equality follows from the complementary slackness property which was proved above. This proves the theorem for $\xi \geq 1$. When $\xi \leq 1$, the result holds by convexity of $m \mapsto m^{1/\xi}$ where the inequality above is reversed.

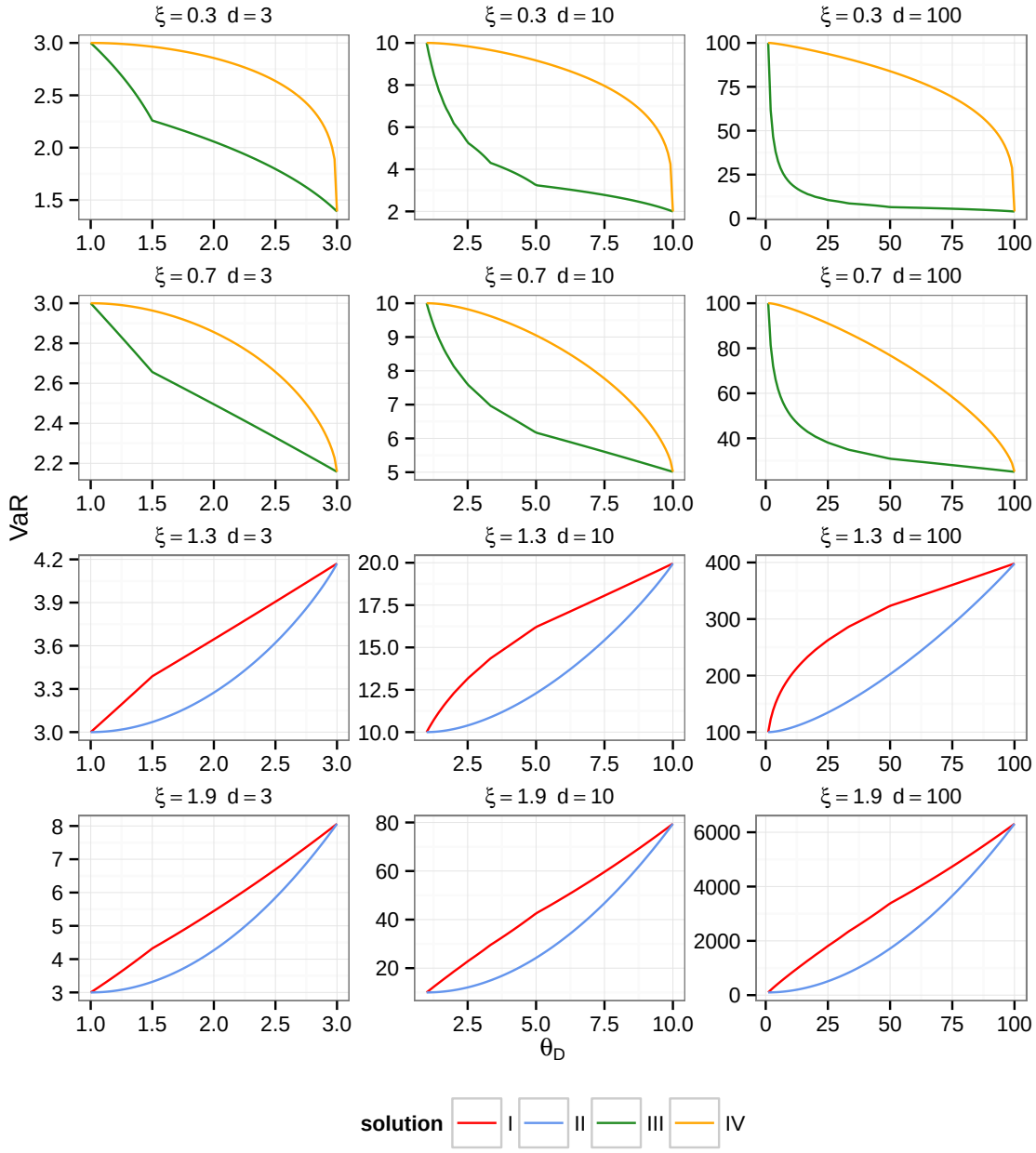
This completes the proof of Theorem 5.14. □

Now we can examine the information on extreme VaR provided by the expressions in (5.17) and (5.18). In this case we are able to compute the information criterion $\mathcal{I}(\boldsymbol{\vartheta})$ for arbitrary dimension and varying shape parameter ξ since negligible computational resource is needed. Results are shown in Table 5.2 and Figure 5.2. Observe that the information provided by a single extremal coefficient increases with the tail index ξ and decreases with dimension d.

5.5 Discussion

In this Chapter we established universal bounds on asymptotic Value-at-Risk for the sum of dependent losses, when full or partial information about extremal coefficients are given. We characterize the bounds as solutions to linear semi-infinite programs and prove that the bounds are obtained by discrete spectral measures supported on a finite number atoms. We illustrate that significant information about extreme Value-at-Risk is obtainable from the readily estimable parameters and provide a partial algorithm for calculating sharp bounds of the Value-at-Risk for the sum of tail dependent losses. Estimation for high dimensional models ($d \geq 16$) is still a

Figure 5.2: Upper and lower bounds on extreme VaR ρ^{ξ} when given a single fixed d-vivariate extremal coefficient ϑ_D constraint.



challenge, nonetheless our experiments suggest that promising results are possible if more sophisticated optimization techniques can be deployed.

Table 5.2: VaR information $\mathcal{I}(\boldsymbol{\vartheta})$ for a single d-variate extremal coefficient.

d	ξ	$\lvert d^{\xi} - d\rvert$	$\min_{i=1,\ldots,n} \mathcal{I}(\boldsymbol{\vartheta}^{(i)})$	$\text{median}_{i=1,\ldots,n} \mathcal{I}(\boldsymbol{\vartheta}^{(i)})$
3	0.3	1.61	0.48	0.52
3	0.7	0.84	0.57	0.63
3	1.3	1.17	0.68	0.75
3	1.9	5.06	0.77	0.82
10	0.3	8.00	0.26	0.36
10	0.7	4.99	0.42	0.54
10	1.3	9.95	0.61	0.71
10	1.9	69.43	0.74	0.81
100	0.3	96.02	0.13	0.23
100	0.7	74.88	0.29	0.43
100	1.3	298.11	0.56	0.67
100	1.9	6209.57	0.73	0.81

5.6 Conic programming.

A *convex cone* is a set $\mathcal{K}$ with the property that if $x, y \in \mathcal{K}$, then $\lambda_1 x + \lambda_2 y \in \mathcal{K}$ for any $\lambda_1, \lambda_2 \geq 0$. A *convex conic program* is a mathematical program that can be formulated in the following way

$$
\begin{aligned}
(CP) \qquad & \inf_{\boldsymbol{x} \in \mathbb{R}^d} && \boldsymbol{c}^{\top}\boldsymbol{x} \\
& \text{subject to:} && \boldsymbol{a}_i^{\top}\boldsymbol{x} = b_i,\ i = 1, \ldots n, \\
& && \boldsymbol{x} \in \mathcal{K}_j,\ j = 1, \ldots, m
\end{aligned}
$$

where $\mathcal{K}_j \subset \mathbb{R}^d,\ j = 1, \ldots, m$ are convex cones. The problem of minimizing a sum of p-norms subject to linear constraints can easily be cast into a conic program. To see

this, fix $p \geq 1$ and consider

$$\begin{aligned} (S) \qquad \inf_{\boldsymbol{x}\in\mathbb{R}^d} \quad & \sum_{k=1}^{q} \|A_k\boldsymbol{x} + c_k\|_p \\ \text{subject to:} \quad & \boldsymbol{a}_i^\top \boldsymbol{x} = b_i,\ i = 1, \dots n, \\ & \boldsymbol{h}_j^\top \boldsymbol{x} \leq d_j,\ j = 1, \dots, m, \end{aligned}$$

which is equivalent to

$$\begin{aligned} (S') \qquad \inf_{(\boldsymbol{x},\boldsymbol{t})\in\mathbb{R}^d\times\mathbb{R}^q} \quad & \sum_{k=1}^{q} t_k \\ \text{subject to:} \quad & \boldsymbol{a}_i^\top \boldsymbol{x} = b_i,\ i = 1, \dots n, \\ & \boldsymbol{h}_j^\top \boldsymbol{x} \leq d_j,\ j = 1, \dots, m, \\ & \|A_k\boldsymbol{x} + c_k\|_p \leq t_k,\ k = 1, \dots, q. \end{aligned}$$

The set $\{(\boldsymbol{x}, \boldsymbol{t}) : \|A_k\boldsymbol{x} + c_k\|_p \leq t_k\}$ is obviously convex and is called a *p-order cone* in $\mathbb{R}^{d+q}$.

While conic programs are convex optimization problems, there does not exist a generic algorithm that computes solutions to general conic programs that scales to high dimension efficiently. A key property of conic constraints that usually leads to an efficient solver is the case where the convex cones $\mathcal{K}_j,\ j = 1, \dots, m$ are *self-dual* (see Nesterov, 2011). Generally this limits the type of programs solved by readily available software to subclasses involving the three known self-dual cones, the non-negative cone $\mathbb{R}^d_+$, the cone of symmetric positive semi-definite matrices and the class of *second-order cone programs* (SOCPs)

$$\begin{aligned} (SOCP) \qquad \inf_{\boldsymbol{x}\in\mathbb{R}^p} \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to:} \quad & \boldsymbol{a}_i^\top \boldsymbol{x} = b_i,\ i = 1, \dots n, \\ & \|A_j^\top \boldsymbol{x} - \boldsymbol{c}_j\|_2 \leq \boldsymbol{h}_j^\top \boldsymbol{x} - d_j,\ j = 1, \dots, m, \end{aligned}$$

where $\mathcal{K}_j = \{\boldsymbol{x} \in \mathbb{R}^p : \|A_j^\top \boldsymbol{x} - \boldsymbol{c}_j\|_2 \leq \boldsymbol{h}_j^\top \boldsymbol{x} - d_j\}$ is a second-order cone. SOCPs appear in a wide variety of applications and numerous solvers exists which can solve SOCPs quickly in relatively high dimension. This implies that the special case of (S') where $p = 2$ can be solved efficiently with readily available software.

5.7 Strong duality, optimality and reducibility of SIPs

For the benefit of the reader, we collect here results from the theory of semi-infinite programming regarding strong duality, optimality conditions, and the reducibility SIPs. These are direct results from the literature and are mostly presented without proof. We carry over all notations and conventions from Section 5.2. The first result gives conditions for strong duality of (P, D).

Proposition 5.15 (Thm. 2.3 in Shapiro, 2009)**.** *If Assumption 5.6 holds and* $\operatorname{val}(P)$ *is finite, then* $\operatorname{sol}(D)$ *is non-empty and* $\operatorname{val}(P) = \operatorname{val}(D)$.

The following proposition establishes versions of the classic Karush–Kuhn–Tucker (KKT) optimality conditions for the case of LSIPs.

Proposition 5.16 (KKT condition)**.** *Suppose Assumption 5.6 is satisfied and* $\operatorname{val}(P)$ *is finite. Fix* $\boldsymbol{x} \in \mathbb{R}^p$*. If there exists* $\boldsymbol{y} \in \mathbb{R}^p_+$ *and* $\{t_1, \ldots, t_p\} \subset T$ *such that*

$$\sum_{k=1}^{p} y_k \boldsymbol{a}(t_k) = \boldsymbol{c}, \tag{5.19}$$

$$\boldsymbol{a}(t_k)^\top \boldsymbol{x} = b(t_k),\ k = 1, \ldots, p, \tag{5.20}$$

and

$$\boldsymbol{a}(t)^\top \boldsymbol{x} \geq b(t),\ \textit{for all } t \in T. \tag{5.21}$$

Then $\boldsymbol{x} \in \operatorname{sol}(P)$.

Note. For arbitrary $\boldsymbol{x} \in \mathbb{R}^p$, define the set of *active indices* $T(\boldsymbol{x}) := \{t \in T : \boldsymbol{a}(t)^\top \boldsymbol{x} = b(t)\}$. By Thm 7.1.(ii) of Goberna and Lopez (1998) (see also Section 11.2 therein), a primal feasible vector $\tilde{\boldsymbol{x}} \in \mathbb{R}^p$ is optimal for (P) if

$$\boldsymbol{c} \in \operatorname{cone}\{\boldsymbol{a}(t) : t \in T(\tilde{\boldsymbol{x}})\}, \tag{5.22}$$

where cone$\{C\}$ denotes the smallest convex cone containing $C \subset \mathbb{R}^p$. Indeed, if there exists $\boldsymbol{y} \in \mathbb{R}^p_+$ and $\{t_1, \ldots, t_p\} \subset T$ such that (5.19) and (5.20) are satisfied, then relation (5.22) holds. □

The final result of this section establishes conditions for the *reducibility* of the LSIP (P).

Proposition 5.17 (Thm. 3.2 in Shapiro, 2009)**.** *Suppose that Assumption 5.6 holds for problem (P) and* val(P) *is finite. If for any $\{t_1, t_2, \ldots t_{p+1}\} \subset T$, there exists $\boldsymbol{x} \in \mathbb{R}^p$ such that*

$$\boldsymbol{a}(t_k)^\top \boldsymbol{x} > b(t_k),\ k = 1, \ldots, p+1.$$

Then there exists $\{t_1, \ldots, t_m\} = T_m \subset T$ with $m \leq p$ such that for corresponding discretizations (P_m) and (D_m)

$$\operatorname{val}(P) = \operatorname{val}(P_m) = \operatorname{val}(D_m) = \operatorname{val}(D).$$

Corollary 5.18. *If Assumption 5.6 and the Slater condition holds for problem (P), then there exists $\omega \in \operatorname{sol}(D) \subset \Omega$ such that ω is finitely supported on at most p atoms $\{t_1, t_2, \ldots, t_p\} \subset T$.*

5.8 Proofs

In this section, denote $2^{\{1,\dots,d\}}$ as the power set of $\{1,\dots,d\}$ and $K^c = \{1,\dots,d\}\backslash K$. To prove Theorem 5.11 we utilize two lemma

Lemma 5.19. *Let $0 \le u_{(1)} \le u_{(2)} \le \cdots \le u_{(d)} \le 1$ be the order statistics for arbitrary $\boldsymbol{u} \in \mathbb{S}_+^{d-1}$. Fix $\xi > 0$ and define $u_{(0)} = 0$. The following equality holds*

$$\sum_{J\in 2^{\{1,\dots,d\}}\backslash\varnothing} \max_{j\in J}\{u_j\} \sum_{L\subset J}(-1)^{|L|+1}|J^c \cup L| = \sum_{j=1}^{d}(d+1-j)^{1/\xi}\left(u_{(j)} - u_{(j-1)}\right).$$

Proof.

$$\begin{aligned}
&\sum_{J\in 2^{\{1,\dots,d\}}\backslash\varnothing} \max_{j\in J}\{u_j\} \sum_{L\subset J}(-1)^{|L|+1}|J^c \cup L| \\
&= \sum_{i=1}^{d} u_{(i)} \left\{ \sum_{J\in 2^{\{1,\dots,d\}}\backslash\varnothing} \mathbb{I}\left(\max_{j\in J}\{u_j\} = u_{(i)}\right) \sum_{k=0}^{|J|} \binom{|J|}{k} (-1)^{k+1}(d-|J|+k)^{1/\xi} \right\} \\
&= \sum_{i=1}^{d} u_{(i)} \left\{ \sum_{\ell=1}^{i} \sum_{\substack{J\in 2^{\{1,\dots,d\}}\backslash\varnothing \\ |J|=\ell}} \mathbb{I}\left(\max_{j\in J}\{u_j\} = u_{(i)}\right) \sum_{k=0}^{|J|} \binom{|J|}{k} (-1)^{k+1}(d-|J|+k)^{1/\xi} \right\} \\
&= \sum_{i=1}^{d} u_{(i)} \left\{ \sum_{\ell=1}^{i} \binom{i-1}{\ell-1} \sum_{k=0}^{\ell} \binom{\ell}{k} (-1)^{k+1}(d-(\ell-k))^{1/\xi} \right\}. \quad (5.23)
\end{aligned}$$

Now fix $i \in \{1,\dots,d\}$ and consider

$$\begin{aligned}
&\sum_{\ell=1}^{i} \binom{i-1}{\ell-1} \sum_{k=0}^{\ell} \binom{\ell}{k} (-1)^{k+1}(d-(\ell-k))^{1/\xi} \\
&= d^{1/\xi} \sum_{k=1}^{i} \binom{i-1}{k-1} (-1)^{k+1} + \sum_{q=1}^{i} (d-q)^{1/\xi} \sum_{k=0}^{i-q} \binom{i-1}{q+k-1} \binom{q+k}{k} (-1)^{k+1} \\
&= d^{1/\xi}\mathbb{I}(i=1) + \sum_{q=1}^{i} (d-q)^{1/\xi}(-1)^{i-q+1}\mathbb{I}(i-q\le 1) = (d+1-i)^{1/\xi} - (d-i)^{1/\xi}.
\end{aligned} \quad (5.24)$$

The first equality in (5.24) holds by summing over all k such that $q = \ell - k,\ q = 0, 1, \ldots, i$. Substituting (5.24) into (5.23) gives

$$\sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \max_{j\in J}\{u_j\} \sum_{L\subset J}(-1)^{|L|+1}|J^c \cup L| = \sum_{i=1}^{d} u_{(i)} \left((d+1-i)^{1/\xi} - (d-i)^{1/\xi}\right)$$
$$= \sum_{i=1}^{d}(d+1-i)^{1/\xi}\left(u_{(i)} - u_{(i-1)}\right).$$

This proves Lemma 5.19. □

The next Lemma establishes analytical solutions to the following special cases of Problems (I) and (III), where the entire set of extremal coefficients take the values $\boldsymbol{\vartheta} = (\vartheta_J)_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \in \mathbb{R}_+^{2^d-1}$

$$(\mathrm{I}'_{\boldsymbol{\vartheta}}) \qquad \inf_{\boldsymbol{x}\in\mathbb{R}^p} \quad \boldsymbol{\vartheta}^\top \boldsymbol{x}$$
$$\text{subject to:} \quad \left(u_1^\xi + u_2^\xi + \cdots + u_d^\xi\right)^{1/\xi} \le \sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \max_{j\in J}\{u_j\} x_J,\ \boldsymbol{u} \in \mathbb{S}_+^{d-1}.$$

$$(\mathrm{III}'_{\boldsymbol{\vartheta}}) \qquad \sup_{\boldsymbol{x}\in\mathbb{R}^p} \quad \boldsymbol{\vartheta}^\top \boldsymbol{x}$$
$$\text{subject to:} \quad \left(u_1^\xi + u_2^\xi + \cdots + u_d^\xi\right)^{1/\xi} \ge \sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \max_{j\in J}\{u_j\} x_J,\ \boldsymbol{u} \in \mathbb{S}_+^{d-1}.$$

Lemma 5.20. *The vector* $\tilde{\boldsymbol{x}} = (\tilde{x}_J)_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing}$ *with elements*

$$\tilde{x}_J := \sum_{L\subset J}(-1)^{|L|+1}|J^c \cup L|^{1/\xi}$$

is optimal for Problems $(\mathrm{I}'_{\boldsymbol{\vartheta}})$ *and* $(\mathrm{III}'_{\boldsymbol{\vartheta}})$ *with*

$$\sum_{K\in 2^{\{1,\dots,d\}}\setminus\varnothing} |K|^{1/\xi}\beta_K = \begin{cases} \mathrm{val}(\mathrm{I}'_{\boldsymbol{\vartheta}}) & \xi \geq 1 \\ \mathrm{val}(\mathrm{III}'_{\boldsymbol{\vartheta}}) & \xi \leq 1, \end{cases}$$

where $(\beta_K)_{K\in 2^{\{1,\dots,d\}}\setminus\varnothing} \in \mathbb{R}^p_+$ *is the unique solution to*

$$\sum_{K\in 2^{\{1,\dots,d\}}\setminus\varnothing} \mathbb{I}\left\{(J\cap K)\neq\varnothing\right\}\beta_K = \vartheta_J,\ J\in 2^{\{1,\dots,d\}}\setminus\varnothing. \tag{5.25}$$

Proof. Fix $p = 2^d-1$. We first prove $\tilde{\boldsymbol{x}} \in \mathrm{sol}(\mathrm{I}'_{\boldsymbol{\vartheta}})$ by verifying the optimality conditions of Proposition 5.16. Thus, we need to show there exists $(y_K)_{K\in 2^{\{1,\dots,d\}}\setminus\varnothing} \in \mathbb{R}^p_+$ and $\{\boldsymbol{u}_K\}_{K\in 2^{\{1,\dots,d\}}\setminus\varnothing} \subset \mathbb{S}^{d-1}_+$ such that the following conditions hold

$$\sum_{K\in 2^{\{1,\dots,d\}}} \max_{j\in J}\{u_{jK}\}y_K = \vartheta_J,\ J\in 2^{\{1,\dots,d\}}\setminus\varnothing, \tag{5.26}$$

$$\sum_{J\in 2^{\{1,\dots,d\}}\setminus\varnothing} \max_{j\in J}\{u_{jK}\}x_J = \left(u_{1K}^\xi + u_{2K}^\xi + \cdots + u_{dK}^\xi\right)^{1/\xi},\ K\in 2^{\{1,\dots,d\}}\setminus\varnothing, \tag{5.27}$$

and

$$\left(u_1^\xi + u_2^\xi + \cdots + u_d^\xi\right)^{1/\xi} \leq \sum_{J\in 2^{\{1,\dots,d\}}\setminus\varnothing} \max_{j\in J}\{u_j\}x_J,\ \text{for all } \boldsymbol{u}\in\mathbb{S}^{d-1}_+. \tag{5.28}$$

Thm. 4 of (Schlather and Tawn, 2002) asserts that β_K in (5.25) always exists for a consistent set of extremal coefficients. Now let $y_K = |K|\beta_K$ and $\boldsymbol{u}_K = |K|^{-1}(\mathbf{1}_J(i))_{i=1}^d \in \mathbb{S}^{d-1}_+$. We verify (5.26)-(5.28) directly.

Condition (5.26):

$$\sum_{K\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \max_{j\in J}\{u_{jK}\}y_K = \sum_{K\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \max_{j\in J}\{|K|^{-1}\mathbf{1}_K(j)\}|K|\beta_K$$
$$= \sum_{K\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \mathbb{I}\left\{(J\cap K)\neq\varnothing\right\}\beta_K = \vartheta_J.$$

Condition (5.27):

$$\sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \max_{j\in J}\{u_{jK}\}\tilde{x}_J =$$
$$\sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \mathbb{I}\left\{(J\cap K)\neq\varnothing\right\}|K|^{-1}\sum_{L\subset J}(-1)^{|L|+1}|J^c\cup L|^{1/\xi}$$
$$= |K|^{-1}\sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \mathbb{I}\left\{(J\cap K)\neq\varnothing\right\}\sum_{L\subset J}(-1)^{|L|+1}|J^c\cup L|^{1/\xi}$$
$$= |K|^{-1}|K|^{1/\xi} = \left(u_{1K}^{\xi}+u_{2K}^{\xi}+\cdots+u_{dK}^{\xi}\right)^{1/\xi},\ K\in 2^{\{1,\ldots,d\}}\setminus\varnothing,$$

where the third equality above follows from a Mobius inversion formula (see Thm. 4 of Schlather and Tawn, 2002).

Condition (5.28):

For $(u_1,\ldots,u_d)^\top\in\mathbb{S}_+^{d-1}$, define $f_k(j)=\mathbb{I}\{k\le j\}\left(u_{(k)}-u_{(k-1)}\right)$ where

$$0=u_{(0)}\le u_{(1)}\le\cdots\le u_{(d)},$$

are the order statistics of $(0,u_1,\ldots,u_d)$. Then by Minkowski's inequality (see

No. 198 of Hardy et al., 1934)

$$\left(u_1^\xi + u_2^\xi + \cdots + u_d^\xi\right)^{1/\xi} = \left\{\sum_{j=1}^d \left(f_1(j) + f_2(j) + \cdots + f_d(j)\right)^\xi\right\}^{1/\xi}$$
$$\leq \left\{\sum_{j=1}^d f_1^\xi(j)\right\}^{1/\xi} + \left\{\sum_{j=1}^d f_2^\xi(j)\right\}^{1/\xi} + \cdots + \left\{\sum_{j=1}^d f_d^\xi(j)\right\}^{1/\xi}$$
$$= \sum_{j=1}^d (d+1-j)^{1/\xi}\left(u_{(j)} - u_{(j-1)}\right) = \sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \max_{j\in J}\{u_j\} \sum_{L\subset J}(-1)^{|L|+1}|J^c \cup L|^{1/\xi}$$
$$= \sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \max_{j\in J}\{u_j\}\tilde{x}_J \quad (5.29)$$

where the second to last equality above follows from Lemma 5.19.

Hence, $\tilde{\boldsymbol{x}} \in \mathrm{sol}(\mathrm{I}'_{\boldsymbol{\vartheta}})$ and

$$\mathrm{val}(\mathrm{I}'_{\boldsymbol{\vartheta}}) = \boldsymbol{\vartheta}^\top\tilde{\boldsymbol{x}} = \sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \tilde{x}_J \sum_{K\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \mathbb{I}\left\{(J\cap K)\neq\varnothing\right\}\beta_K$$
$$= \sum_{J\in 2^{\{1,\ldots,d\}}\setminus\varnothing}\sum_{K\in 2^{\{1,\ldots,d\}}\setminus\varnothing} \mathbb{I}\left\{(J\cap K)\neq\varnothing\right\}\beta_K \sum_{L\subset J}(-1)^{|L|+1}|J^c\cup L|^{1/\xi}$$
$$= \sum_{K\in 2^{\{1,\ldots,d\}}\setminus\varnothing} |K|^{1/\xi}\beta_K.$$

Thus we have proven the Lemma for $(\mathrm{I}'_{\boldsymbol{\vartheta}})$. Now the same holds for Problem $(\mathrm{III}'_{\boldsymbol{\vartheta}})$ by considering $\sup_{\boldsymbol{x}\in\mathbb{R}^p}\boldsymbol{\vartheta}^\top\boldsymbol{x} = -\inf_{\boldsymbol{x}\in\mathbb{R}^p} -\boldsymbol{\vartheta}^\top\boldsymbol{x}$. In this case $\xi \leq 1$ and No. 198 of Hardy et al. (1934) imply that the inequality in (5.29) is reversed. All other steps remain the same. This completes the proof of Lemma 5.20. □

We now prove Theorem 5.11.

Proof of Theorem 5.11. Let $\mathcal{H}_{\boldsymbol{c}}$ denote the space of finite Borel measures on $\mathbb{S}_+^{d-1}$ satisfying

$$\left\{\int_{\mathbb{S}_+^{d-1}} \max_{j\in J}\{u_j\}H(d\boldsymbol{u}) = c_J\right\}_{J\in\mathcal{J}}.$$

Likewise, denote $\mathcal{H}_{\boldsymbol{\vartheta}}$ as the space of finite Borel measures on $\mathbb{S}_+^{d-1}$ satisfying

$$\left\{ \int_{\mathbb{S}_+^{d-1}} \max_{j\in J}\{u_j\} H(d\boldsymbol{u}) = \vartheta_J \right\}_{J\in 2^{\{1,\dots,d\}}\setminus\varnothing}.$$

Hence, we may write Problem (I) as

$$\begin{aligned}\mathrm{val(I)} = \sup_{H\in\mathcal{H}_{\boldsymbol{c}}} \int_{\mathbb{S}_+^{d-1}} \left(u_1^\xi + \cdots + u_d^\xi\right)^{1/\xi} H(d\boldsymbol{u}) \\ = \sup_{\boldsymbol{\vartheta}\in\Theta_{\boldsymbol{c}}} \left\{ \sup_{H\in\mathcal{H}_{\boldsymbol{\vartheta}}} \int_{\mathbb{S}_+^{d-1}} \left(u_1^\xi + \cdots + u_d^\xi\right)^{1/\xi} H(d\boldsymbol{u}) \right\}, \quad (5.30)\end{aligned}$$

where $\Theta_{\boldsymbol{c}} = \{\boldsymbol{\vartheta}\in\Theta : \vartheta_J = c_J, \text{ for all } J\in\mathcal{J}\}$. (Recall Θ is the space of consistent extremal coefficients). Now Lemma 5.20 together with strong duality for $(\mathrm{I}'_{\boldsymbol{\vartheta}})$ imply

$$\sup_{H\in\mathcal{H}_{\boldsymbol{\vartheta}}} \int_{\mathbb{S}_+^{d-1}} \left(u_1^\xi + \cdots + u_d^\xi\right)^{1/\xi} H(d\boldsymbol{u}) = \mathrm{val}(\mathrm{I}'_{\boldsymbol{\vartheta}}) = \sum_{K\in 2^{\{1,\dots,d\}}\setminus\varnothing} |K|^{1/\xi}\beta_K, \quad (5.31)$$

with $(\beta_K)_{K\in 2^{\{1,\dots,d\}}\setminus\varnothing} \in \mathbb{R}_+^p$ being the unique solution to

$$\sum_{K\in 2^{\{1,\dots,d\}}\setminus\varnothing} \mathbb{I}\{(J\cap K)\neq\varnothing\}\,\beta_K = \vartheta_J,\ J\in 2^{\{1,\dots,d\}}\setminus\varnothing.$$

Substituting (5.31) into (5.30) gives

$$\begin{aligned}\mathrm{val(I)} \;=\; & \sup_{\boldsymbol{\beta}\in\mathbb{R}_+^{2^d-1}} \sum_{J\subset\{1,\dots,d\}\setminus\varnothing} |J|^{1/\xi}\beta_J, \\ & \text{subject to : } \left\{ \sum_{K\subset\{1,\dots,d\}\setminus\varnothing} \mathbb{I}\{(K\cap J)\neq\varnothing\}\,\beta_K = c_J \right\}_{J\in\mathcal{J}}.\end{aligned}$$

This proves the Theorem in the case of Problem (I) where $\xi\geq 1$. For Problem (II) where $\xi\leq 1$, the proof is nearly identical whence replacing sup by inf in (5.30). □

APPENDIX

APPENDIX A

Additional R Code

A.1 Code for fitting max-linear models via CRPS

Below we provide R code for fitting simple max-linear models. The script requires the package `matrixStats` (Bengtsson, 2014) which can be installed from CRAN.

```
# requires matrixStats package
require(matrixStats)

## function maxlinear: same as A %*% Z except replaces plus
## operation by max
maxlinear <- function(A,Z){
  d <- nrow(A)
  p <- ncol(A)
  if(class(Z)=="numeric"){
    Z <- matrix(Z,length(Z),1)
  }
  if(p != nrow(Z)) stop("non-conformable arguments")
  n <- ncol(Z)
  X <- matrix(rowMaxs(A[rep(1:d,n),]*t(Z[,rep(1:n,each = d)])),d,n)
  return(X)
}

## function frechetCRPS calculates the CRPS of a 1-Frechet
## distribution with scale 'sigma' given observations 'obs'
frechetCRPS <- function(sigma, obs){
  q <- 1/obs
  C1 <- 1/(2*sigma + 1)
  C2 <- -2*(1 - exp(-(sigma + 1)*q))/(sigma + 1)
  C3 <- 1 - exp(-q)
  crps <- C1 + C2 + C3
  return(crps)
}

## function maxlinearCRPS: given the coefficient maxtrix 'A'
```

```
## of a max-linear model, observations 'x' and atoms 'U' of the
## CRPS weighting measure.  Returns the CRPS score.
## optionally one can specify 'U' as the number of atoms and
## 'seed' then the weigting measure will be randomly
## generated from the ncol(x) - 1 dimensional simplex with
## number of atoms 'U'.
maxlinearCRPS <- function(A, x, U = 100, seed = 1){
  d <- ncol(x)
  if(nrow(A) != d) stop("A and x non-conformable arguments")
  if(class(U) == "numeric"){
    if(seed >= 0) set.seed(seed)
    U <- t(apply(matrix(runif((d-1)*U),U,d-1),1,function(z){
      z <- diff(c(0,sort(z)))
      return(c(z,1-sum(z)))
    }))
  }
  if(ncol(U) != d) stop("U and x non-conformable arguments")
  n <- nrow(x)
  m <- nrow(U)
  V <- matrix(rowSums(maxlinear(1/U,A)),n,m,byrow = TRUE)
  M <- maxlinear(x, t(1/U))
  crps <- frechetCRPS(V,M)
  return(sum(crps))
}


## function fit.simple.maxlinear.  Fits a simple max-linear model given
## data 'x'. The user must supply the number of factors 'p' in the model.
## currently the only method implemented is the CRPS method.  Optionally
## one can supply the atoms of the weighting measure as 'U'.
fit.simple.maxlinear <- function(x, p = 2, U = 100, seed = 1, start = 0, method = "CRPS",
                                 optim.method = "Nelder-Mead", optim.control = list()){
  d <- ncol(x)
  if(class(U) == "numeric"){
    if(seed >= 0) set.seed(seed)
    U <- t(apply(matrix(runif((d-1)*U),U,d-1),1,function(z){
      z <- diff(c(0,sort(z)))
      return(c(z,1-sum(z)))
    }))
  }
  crpsOptFun <- function(aa,pp,xx,dd,UU){
    A <- matrix(aa,dd,pp-1)
    A <- cbind(A,1 - rowSums(A))
    return(maxlinearCRPS(A,xx,UU))
  }
  ui <- rbind(diag(d*(p-1)),-diag(d)[,rep(1:d,p-1)])
  ci <- c(rep(0,d*(p-1)),rep(-1, d))
  if(start >= 0) set.seed(start)
  A.start <- matrix(runif(d*p),d,p)
  A.start <- A.start/matrix(rep(rowSums(A.start),p),d,p)
  a.start <- c(A.start[,-p])
  copt <- constrOptim(theta = a.start, f = crpsOptFun, grad = NULL,
                      ui = ui, ci = ci,
                      pp = p, xx = x, dd = d, UU = U)
  A.opt <- matrix(copt$par,d,p-1)
  A.opt <- cbind(A.opt, 1- rowSums(A.opt))
  return(list(A = A.opt, optim = copt))
}

## Examples

n <- 100 # number observations
d <- 4   # model dimension
p <- 3   # number of factors in max-linear model
m <- 1000 # number of atoms in CRPS weighting measure

set.seed(12)
A <- matrix(runif(d*p),d,p)
```

```
A <-  diag(1/rowSums(A)) %*% A
Z <- matrix(1/rexp(p*n),p,n)
x <- t(maxlinear(A,Z))

mlfit <- fit.simple.maxlinear(x = x, p = p, U = m, optim.control = list(maxit = 500))

maxlinearCRPS(mlfit$A,x)
maxlinearCRPS(A,x)
```

A.2 Code for fitting a Tawn-Molchanov max-stable model

Here we provide documentation on `R` code written for fitting TM models as described in Chaper 4. After downloading the software from Yuen (2014), one needs install the `R` packages `matrixStats`, `quadprog` and `evd`. Finally, source the file `tm_model_functions.R`. This will make available the function `fitTM` and its dependencies. This function fits the Tawn-Molchanov max-stable model to data `x` - realizations from a max-stable random vector. Optionally, it estimates Value-at-Risk for a given quantile using empirical samples from the fitted model.

```
Fit a Tawn-Molchanov max-stable model

Description:

     This function fits the Tawn-Molchanov max-stable model to data x
     assumed to be realizations from a max-stable random vector.
     Optionally, estimate Value-at-Risk for a given quantile using an
     emperical sample from the fitted model.

Usage:

     fitTM(x, marginal.gev = list(fit = FALSE, pooled = FALSE, loc = 1, scale = 1, shape = 1),
       bootstrap.se = FALSE, est.VaR = list(n = 0, alpha = 0.95, weights = 1, func = "sum",
       block.size = 1, bagged = list(m = 100, quantile = 0.5)),
       multcore = FALSE)

Arguments:

       x: A data matrix to fit the model to. Each row corresponds to an
          independent realization.  Columns correspond to components of
          the model.

marginal.gev: A list containing the following elements: -fit a
          logical indicating whether to fit the marginal GEV parameters
          using maximum likelihood. The default is FALSE.  -pooled
          a logical indicating whether to treat the columns of x as
          identically distributed.  If FALSE (the default) GEV
          parameters location, shape and scale will be fit for
          each column of x. If TRUE, all columns of x are pooled
          and a single location, shape and scale parameter is
          fitted.  -loc A numeric scalar or vector the same length as
          ncol(x).  If specified the location parameter(s) is fixed
          to the given value.  -scale A numeric scalar or vector the
```

```
          same length as ncol(x).  If specified the scale
          parameter(s) is fixed to the given value.  -shape A numeric
          scalar or vector the same length as ncol(x).  If specified
          the shape parameter(s) is fixed to the given value.

bootstrap.se: A numeric indicating the number of bootstrap samples that
          should be used to estimate standard errors for all
          parameters.  Setting to 0 or FALSE (the default) does not
          perform bootstrapping.

 est.VaR: A list containing the following elements: -n The size of
          the emperical sample used to estimate Value-at-Risk (VaR).
          If 0 no estimation of VaR will be performed.  -alpha The
          quantile of Value-at-Risk to be estimated (usually alpha is
          close to 1) -weights A vector the same length as ncol(x)
          indicating the weight of each component of x in VaR.
          -func A string indicating the functional for which
          Value-at-Risk should be estimated.  Current valid functionals
          are "sum" for the sum or "max" for the maximum.
          -block.size Depricated, should always be set to 1.
          -bagged A list indicating if the empirical estimating
          procedure for Value-at-Risk should be bagged to reduce
          variance of the estimate.  Must contain m, a numeric scalar
          indicating the size of the bag and quantile, either a
          numeric value between zero and one indicating which quantaile
          of the bagged sample to use as the estimate (the default is
          the median 0.5) or the character string "mean" which will
          use the average of the bagged sample as the estimate for VaR.
          If m is zero, no bagging is performed.

multcore: Make use of multipule cores? - Not yet implemented.

Value:

     A list containing the following elements

beta.hat: The fitted Tawn-Molchanov coefficients of the model.

 ecf.mle: Naive estimates of the extremal coefficients used in the
          quadratic program.

      qp: Additional information returned from quadprog.

loc, scale, shape: If marginal.gev is specified, the fitted marginal
          GEV parameters.

     VaR: If est.VaR is specified, the estimated Value-at-Risk at the
          given alpha quantile.

boostrap: If bootstrap.se is specied, the bootstrap standard errors
          for each parameter fitted.

Author:

     R.A. Yuen

Examples:

       d <- 5
       n <- 100
       A <- matrix(c(1,0,0,0,0,0,
                0.5,0,0.5,0,0,0,
                rep(1/6,6),
                0.5,rep(0.1,5),
                0,0,0,rep(1/3,3)),d,6, byrow = TRUE)
       p <- ncol(A)
       Z <- matrix(1/rexp(p*n),p,n)
       x <- t(maxlinear(A,Z))
```

```
simple.tm.fit <- fitTM(x = x)


shp <- 0.5
scl <- seq(0.5,2, length.out = d)
v <- frech2gev(x, loc = 1, scale = scl, shape = shp)
tm.fit <- fitTM(x = v, marginal.gev = list(fit = TRUE, pooled = FALSE))
```

BIBLIOGRAPHY

BIBLIOGRAPHY

Balkema, A. and Resnick, S. (1977). Max-infinite devisibility. *Journal of Applied Probability*, 14(2):309–319.

Bank for International Settlements (2011). *Basel III: A global regulatory framework for more resilient banks and banking systems*. `http://www.bis.org/publ/bcbs189.pdf`.

Barbe, P., Fougeres, A., and Genest, C. (2006). On the tail behavior of sums of dependent risks. *Astin Bull*, 36:361–373.

Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of extremes*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester. Theory and applications, With contributions from Daniel De Waal and Chris Ferro.

Bengtsson, H. (2014). *matrixStats: Methods that apply to rows and columns of a matrix*. R package version 0.8.14.

Berrocal, V., Gelfand, A., and Holland, D. (2014). Assessing exceedance of ozone standards: a space-time downscaler for fourth highest ozone concentrations. *Environmetrics*, 25(4):279–291.

Bingham, N. H., Goldie, C. M., and Teugels, J. L. (1987). *Regular Variation*. Cambridge University Press.

Brown, B. M. and Resnick, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Probability*, 14(4):732–739.

Coles, S. (2001). *An Introduction to the Statistical Modeling of Extreme Values*. Springer.

Coles, S., Heffernan, J., and Tawn, J. A. (1999). Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365.

Coles, S. G. and Tawn, J. A. (1996). Modelling extremes of the areal rainfall process. *Journal of the Royal Statistical Society, Series B*, 58(2):329–347.

Cooley, D., Naveau, P., and Poncet, P. (2006). Variograms for spatial max-stable random fields. In *Dependence in probability and statistics*, volume 187 of *Lecture Notes in Statist.*, pages 373–390. Springer, New York.

Davison, A. and Blanchet, J. (2011). Spatial modeling of extreme snow depth. *Annals of Applied Statistics*, 5(3):1699–1725.

Davison, A., Padoan, S., and Ribatet, M. (2012). The statistical modeling of spatial extremes. *Statistical Science*.

de Haan, L. (1978). A characterization of multidimensional extreme-value distributions. *Sankhyā (Statistics). The Indian Journal of Statistics. Series A*, 40(1):85–88.

de Haan, L. (1984). A spectral representation for max-stable processes. *Annals of Probability*, 12(4):1194–1204.

de Haan, L. and Ferreira, A. (2006). *Extreme value theory*. Springer Series in Operations Research and Financial Engineering. Springer, New York. An introduction.

DeOliveira, V. (2005). Bayesian inference and prediction of gaussian random fields based on censored data. *Journal of Computational and Graphical Statistics*, 14(1):95–115.

Dombry, C., Eyi-Minko, F., and Ribatet, M. (2012). Conditional simulation of max-stable processes. *Biometrika*.

Dupacova, J. and Wets, R. (1988). Asymptotic behavior of statistical estimators and of optimal solutions to stochastic optimization problems. *The Annals of Statistics*, 16:1517–1549.

Einmahl, J. H., Krajina, A., and Segers, J. (2012). An M-estimator for tail dependence in arbitrary dimension. *Annals of Statistics*, 40(3):1764–1793.

Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extreme Events*. Springer-Verlag, New York.

Embrechts, P., Lambrigger, D. D., and Wuthrich, M. V. (2009). Multivariate extremes and the aggregation of dependent risks: examples and counter-examples. *Extremes*, 12:107–127.

Embrechts, P., Puccetti, G., Rüschendorf, L., and Wang, R. (2014). An academic response to Basel 3.5. *Risks*, 2:25–48.

Engelke, S., Malinowski, A., Kabluchko, Z., and Schlather, M. (2012). M-estimation of husler-reiss distributions and brown-resnick processes. *arXiv:1207.6886v2*.

Erhardt, R. and Smith, R. (2012). Approximate bayesian computing for spatial extremes. *Computational Statistics and Data Analysis*, 56(6):1468–1481.

Ferreira, A. and de Haan, L. (2014). The generalized pareto process; with a view towards application and simulation. *Bernoulli*, 20(4):1717–1737.

Finkenstädt, B. and Rootzén, H., editors (2004). *Extreme Values in Finance, Telecommunications, and the Environment*, volume 99 of *Monographs on Statistics and Applied Probability*. Chapman and Hall / CRC, New York.

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*, 28:180.

Fougères, A.-L., Mercadier, C., and Nolan, J. P. (2013). Dense classes of multivariate extreme value distributions. *J. Multivariate Analysis*, 116:109–129.

Fougères, A.-L., Nolan, J. P., and Rootzén, H. (2009). Models for dependent extremes using stable mixtures. *Scand. J. Stat.*, 36(1):42–59.

French, K. (2014). Data library of standardized industry portfolios. Available online: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient Metropolis jumping rules. In *Bayesian statistics, 5 (Alicante, 1994)*, Oxford Sci. Publ., pages 599–607. Oxford Univ. Press, New York.

Gnedenko, B. V. (1941). Limit theorems for the maximal term of a variational series. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 32:7–9.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268.

Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Goberna, M. and Lopez, M. (1998). *Linear Semi-infinite Optimization*. Wiley series in mathematical methods in practice. Wiley.

Gurobi Optimization, I. (2015). Gurobi optimizer reference manual.

Hamidieh, K., Stoev, S., and Michailidis, G. (2009). On the estimation of the extremal index based on scaling and resampling. *J. Comput. Graph. Statist.*, 18(3):731–755.

Hardy, G., Littlewood, J., and Pólya, G. (1934). *Inequalities*. Cambridge University Press.

Kabluchko, Z., Schlather, M., and de Haan, L. (2009). Stationary max–stable fields associated to negative definite functions. *Annals of Probability*, 37(5):2042–2065.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York.

Molchanov, I. (2008). Convex geometry of max-stable distributions. *Extremes*, 11:235–259.

Nesterov, U. (2011). Towards non-symmetric conic optimization. *Optimization methods and software*, 27.

Oesting, M., Kabluchko, Z., and Schlather, M. (2011). Simulation of brown-resnick processes. *Extremes*.

Padoan, S., Ribatet, M., and Sisson, S. (2010). Likelihood-based inference for max-stable processes. *Journal of the American Statistical Association*, 105(489):263–277.

Parr, W. and Schucany, W. R. (1980). Minimum distance and robust estimation. *Journal of the American Statistical Association*, 75(371):616–624.

Reich, B. and Shaby, B. (2012). A finite-dimensional construction of a max-stable process for spatial extremes. *Annals of Applied Statistics*, 6(4).

Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York.

Resnick, S. I. (2007). *Heavy-tail phenomena*. Springer Series in Operations Research and Financial Engineering. Springer, New York. Probabilistic and statistical modeling.

Ribatet, M. (2011). *SpatialExtremes: Modelling Spatial Extremes*. R package version 1.8-1.

Sang, H. and Gelfand, A. (2009). Continuous spatial process models for spatial extreme values. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:49–65.

Schlather, M. (2002). Models for stationary max-stable random fields. *Extremes*, 5(1):33–44.

Schlather, M. and Tawn, J. A. (2002). Inequalities for the extremal coefficients of multivariate extreme value distributions. *Extremes*, 5(1):87–102.

Shaked, M. and Shanthikumar, J. (2007). *Stochastic Orders*. Springer.

Shapiro, A. (2009). Semi-infinite programming, duality, discretization and optimality conditions. *Optimization*, 58(2):133–161.

Smith, R. (1990). Max-stable processes and spatial extremes,. `http://www.unc.edu/depts/statistics/postscript/rs/spatex.pdf`.

Smith, R. (1994). Multivariate threshold methods. In Galambos, J., Lechner, J., and Simiu, E., editors, *Extreme vault theory and applications*, pages 225–248. Kluwer.

Stephenson, A. G. (2002). evd: Extreme value distributions. *R News*, 2(2):0.

Stoev, S. and Taqqu, M. S. (2005). Extremal stochastic integrals: a parallel between max–stable processes and $\alpha-$stable processes. *Extremes*, 8:237–266.

Stoev, S. A. (2008). On the ergodicity and mixing of max-stable processes. *Stochastic Process. Appl.*, 118(9):1679–1705.

Strokorb, K. and Schlather, M. (2013). An exceptional max-stable process fully parameterized by its extremal coefficients. *Bernoulli*. to appear.

Szekely, G. J. and Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, 93(1):58–80.

Thibaud, E., Mutzner, R., and Davison, A. (2013). Threshold modeling of extreme spatial rainfall. *Water resources research*, 49:4633–4644.

Thibaud, E. and Opitz, T. (2013). Efficient inference and simulation for elliptical pareto processes. *arXiv:1401.0168v1*.

Turlach, B. and Weingessel, A. (2013). *quadprog: Functions to solve Quadratic Programming Problems.* R package version 1.5-5.

van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wang, Y. and Stoev, S. A. (2011). Conditional sampling for spectrally discrete max-stable random fields. *Adv. in Appl. Probab.*, 43(2):461–483.

Wolfowitz, J. (1957). The minimum distance method. *Annals of Mathematical Statistics*, 28(1):75–88.

Xue, G. and Ye, Y. (2000). An efficient algorithm for minimizing a sum of p-norms. *SIAM J. Optim*, 10(2):551–579.

Yuen, R. (2014). R-code for fitting the Tawn–Molchanov max-stable model and computing bounds on VaR–max. Available online: `http://hdl.handle.net/2027.42/107443`.